



Institute for Cell and Molecular
Biosciences

The role of SMN in box C/D snoRNP biogenesis

Charles Maurice Debieux
B.Sc(Hons), M.Res,

Submitted for Philosophiae Doctor degree
Faculty of Medical Science
Newcastle University
June 2009

Declaration

I certify that this thesis contains my own work, except where acknowledged, and that no part of this material has been previously submitted for a degree or any other qualification at this or any other University.

Abstract

Box C/D small nucleolar ribonucleic protein complexes (snoRNPs) either direct 2'-*O*-methylation of the pre-ribosomal RNA (pre-rRNA) or function as chaperones in pre-rRNA processing. It has been demonstrated that box C/D snoRNP biogenesis is a highly intricate and co-ordinated process that is mediated by a large, dynamic complex, known as the pre-snoRNP. This complex contains the box C/D snoRNA and a number of common core proteins, which become stably associated with the snoRNA, as well as factors linked to assembly, 3' RNA processing and transport. Interestingly, the product of the gene linked to the neurodegenerative disorder spinal muscular atrophy, the survival of motor neurons protein (SMN), interacts with the box C/D snoRNP common core protein fibrillarin and is vital for the cellular localisation of both fibrillarin and box C/D snoRNA. The SMN protein operates with Gemins2-8 and UNRIP in what is termed the SMN complex, which has known functions in the assembly of the spliceosomal small nuclear RNPs (snRNPs). As the SMN protein interacts with fibrillarin and is also required for the cellular localisation of both fibrillarin and box C/D snoRNA this study set out to investigate the association of fibrillarin with the box C/D snoRNPs and the role of the SMN complex in this process.

In this study it was revealed that the C terminal domain of fibrillarin is essential for cellular localisation and interactions with the box C/D snoRNP assembly factors, which suggests that this domain may mediate fibrillarin association with the box C/D snoRNPs. Also in this study the SMN protein was shown to interact with the box C/D snoRNP assembly factors NUFIP and BCD1, which are stable components of the pre-snoRNP. Analysis of the interaction of SMN with the pre-snoRNP, however, indicates that if SMN does associate with the pre-snoRNP then it is only a transient interaction. Further analysis revealed that as well as the SMN protein, Gemin2, 5, 6 and 7 are essential for the localisation of box C/D snoRNA and that Gemin5 is also required for the accumulation of box C/D snoRNA. This study strengthens the case that the SMN complex is involved in box C/D snoRNP biogenesis, with the data suggesting that it functions as a transport factor rather than in the association of fibrillarin with the box C/D snoRNPs. Analysis of the snRNP transport factor, snurportin1, revealed that it also interacts with numerous components of the pre-snoRNP; however, does not interact with SMN.

Acknowledgments

I would like to thank my supervisor Dr. Nick Watkins for his help throughout the course of this PhD as well as my co-supervisor Dr. Jeremy Brown. I would also like to acknowledge the support given to me from the whole Watkins lab, namely, Dr. Kenneth McKeegan, Dr. Hannah Richardson and fellow PhD students Amy Turner, Andrew Knox and Kate Sloan. Thanks to everyone at home for their support and my stepfather Andrew and family friend Kelly for proof reading this thesis.

This work was funded by the BBSRC.



Contents

| | |
|------------------------------|-------------|
| Declaration..... | I |
| Abstract..... | II |
| Acknowledgements..... | III |
| Contents..... | IV |
| List of Figures..... | X |
| List of Tables..... | XV |
| Abbreviations..... | XVII |

| | |
|---------------------------------------|----------|
| Chapter one: Introduction..... | 1 |
|---------------------------------------|----------|

| | |
|---|----|
| 1.1. Non-coding RNAs (ncRNAs)..... | 2 |
| 1.2. Small nucleolar ribonucleic protein complexes (snoRNPs)..... | 4 |
| 1.2.1. Box C/D snoRNPs..... | 5 |
| 1.2.2. Box H/ACA snoRNPs..... | 8 |
| 1.2.3. The small Cajal body-specific RNPs (scaRNPs)..... | 10 |
| 1.3. Ribosome biogenesis..... | 12 |
| 1.3.1. rDNA gene organisation and transcription..... | 12 |
| 1.3.2. Processing of the large pre-rRNA transcript..... | 13 |
| 1.3.3. Chemical modification of rRNA..... | 16 |
| 1.3.4. A number of snoRNPs function as chaperones in pre-rRNA processing..... | 17 |
| 1.3.5. Ribosomal assembly and export..... | 20 |
| 1.4. Box C/D snoRNP biogenesis..... | 21 |
| 1.4.1. Gene organisation and transcription..... | 21 |
| 1.4.2. Processing of box C/D snoRNAs..... | 22 |
| 1.4.3. 5' cap hypermethylation of box C/D snoRNAs..... | 26 |
| 1.4.4. Role of assembly factors in box C/D snoRNP biogenesis..... | 28 |
| 1.4.4.1 TIP48 and TIP49..... | 29 |
| 1.4.4.2. BCD1..... | 30 |
| 1.4.4.3. NOP17..... | 31 |
| 1.4.4.4. NUFIP..... | 31 |
| 1.4.4.5. TAF9..... | 32 |
| 1.4.5. Transport factors..... | 32 |

| | |
|--|-----------|
| 1.4.5.1. Nuclear export factors..... | 32 |
| 1.4.5.2. Snurportin1..... | 33 |
| 1.4.5.3. Nopp140..... | 34 |
| 1.4.6. Sites of box C/D snoRNP biogenesis..... | 34 |
| 1.5. snRNPs..... | 36 |
| 1.5.1. snRNP structure..... | 36 |
| 1.5.2. Pre-mRNA splicing and snRNPs..... | 38 |
| 1.5.3. snRNP Biogenesis..... | 40 |
| 1.5.3.1. Transcription and nuclear export..... | 40 |
| 1.5.3.2. Cytoplasmic modifications..... | 40 |
| 1.5.3.3. Nuclear import..... | 42 |
| 1.5.3.4. Nuclear maturation..... | 44 |
| 1.5.4. U6 snRNP biogenesis..... | 44 |
| 1.6. SMN and spinal muscular atrophy (SMA)..... | 45 |
| 1.6.1. The SMN complex..... | 45 |
| 1.6.2. Genetics of SMA..... | 47 |
| 1.6.3. SMN, SMA and snRNP biogenesis..... | 48 |
| 1.6.4. SMN, SMA and mRNA trafficking..... | 49 |
| 1.6.5. SMN, SMA and muscle specific functions..... | 50 |
| 1.7. SMN and box C/D snoRNP biogenesis..... | 51 |
| 1.8. Sub nuclear structures | 52 |
| 1.8.1. The nucleolus..... | 52 |
| 1.8.2. Cajal bodies..... | 54 |
| 1.8.3. Nuclear Gems..... | 55 |
| 1.8.4. Nuclear speckles..... | 55 |
| 1.8.5. OPT (oct1 / PTF / transcription) domains..... | 55 |
| 1.8.6. Cleavage bodies..... | 55 |
| 1.8.7. Perinucleolar compartment..... | 56 |
| 1.8.8. PML (promyelocytic leukaemia) bodies..... | 56 |
| 1.9. Research aims..... | 57 |
| Chapter two: Materials and methods..... | 58 |
| 2.1. General molecular biology..... | 59 |

| | |
|--|----|
| 2.1.1. Polymerase chain reaction (PCR)..... | 59 |
| 2.1.2. Agarose gel electrophoresis..... | 59 |
| 2.1.3. DNA extraction and purification from agarose gel..... | 59 |
| 2.1.4. DNA quantification..... | 59 |
| 2.1.5. Restriction enzyme digestion..... | 60 |
| 2.1.6. Ligation of isolated DNA into plasmid..... | 60 |
| 2.1.7. Transformation of bacteria with plasmids..... | 60 |
| 2.1.8. DNA sequencing..... | 60 |
| 2.2. DNA constructs..... | 61 |
| 2.2.1. U3 and U14 snoRNA pcDNA5 expression plasmids..... | 61 |
| 2.2.2. Fibrillarin and SMN pcDNA5 expression plasmids..... | 62 |
| 2.2.3. Degradation domain (DD) proteotuner plasmid..... | 63 |
| 2.2.4. Fibrillarin deletion mutants..... | 63 |
| 2.3. Cell culture, transfection and extract preparation..... | 64 |
| 2.3.1. Cell culture..... | 64 |
| 2.3.2. Development of inducible HEK293 cells..... | 65 |
| 2.3.3. Electroporation..... | 67 |
| 2.3.4. Chemical transfection of plasmids..... | 67 |
| 2.3.5. siRNA mediated depletion..... | 67 |
| 2.3.6. Cell extract preparation..... | 68 |
| 2.3.7. Glycerol gradient preparation..... | 69 |
| 2.3.8. Immunoprecipitation (IP)..... | 70 |
| 2.4. Protein separation and analysis..... | 70 |
| 2.4.1. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS- PAGE)..... | 70 |
| 2.4.2. Coomassie blue staining..... | 71 |
| 2.4.3. Western blot assays..... | 71 |
| 2.5. RNA extraction, separation and analysis..... | 72 |
| 2.5.1. RNA extraction and separation..... | 72 |
| 2.5.2. Northern blot hybridisation..... | 73 |
| 2.5.3. Primer extension assays..... | 74 |
| 2.6. Microscopy..... | 75 |
| 2.6.1. FISH..... | 75 |
| 2.6.2. Immunofluorescence..... | 76 |

| | |
|---|----|
| 2.6.3. Widefield microscopy..... | 77 |
| 2.7. Protein expression, purification and <i>in vitro</i> interaction assays..... | 77 |
| 2.7.1. Recombinant protein expression and purification..... | 77 |
| 2.7.2. Expression of <i>in vitro</i> -translated [³⁵ S] methionine-labelled proteins..... | 79 |
| 2.7.3. <i>In vitro</i> protein-protein interaction assays..... | 80 |
| 2.7.4. <i>In vitro</i> protein-protein interaction assays using thioredoxin tagged proteins..... | 81 |

Chapter three: Development of approaches to analyse Box C/D snoRNP biogenesis.....82

| | |
|---|-----|
| 3.1. Introduction..... | 83 |
| 3.2. Results..... | 87 |
| 3.2.1. Role of the common core proteins in box C/D snoRNP localisation..... | 87 |
| 3.2.2. Loss of the unrelated protein, hPRP31, does not effect box C/D snoRNA localisation..... | 91 |
| 3.2.3. Localisation of the box C/D snoRNP assembly factors..... | 93 |
| 3.2.4. Development of procedures to analyse <i>de novo</i> box C/D snoRNP biogenesis in HeLa SS6 cells..... | 97 |
| 3.2.5. Development of inducible box C/D snoRNA cell lines..... | 101 |
| 3.2.6. Inducible protein systems..... | 105 |
| 3.3. Discussion..... | 124 |

Chapter four: Analysis of the localisation and interactions of fibrillarin130

| | |
|---|-----|
| 4.1. Introduction..... | 131 |
| 4.2. Results..... | 136 |
| 4.2.1. Construction of human fibrillarin deletion mutants..... | 136 |
| 4.2.2. Analysis of the localisation and association, with the box C/D snoRNPs, of the fibrillarin deletion mutants..... | 137 |
| 4.2.3. Fibrillarin interacts directly with the box C/D snoRNP assembly factors..... | 140 |
| 4.2.4. The box C/D snoRNP assembly factors interact with the C terminal domain of fibrillarin..... | 143 |

| | |
|----------------------|-----|
| 4.3. Discussion..... | 146 |
|----------------------|-----|

Chapter five: The role of the SMN complex in box C/D snoRNP biogenesis.....149

| | |
|---|-----|
| 5.1. Introduction..... | 150 |
| 5.2. Results..... | 155 |
| 5.2.1. Fibrillarin interacts with SMN but not with any of the other components of the SMN complex..... | 155 |
| 5.2.2. SMN interacts with the box C/D snoRNP assembly factors BCD1 and NUFIP..... | 157 |
| 5.2.3. There is an overlap in the sedimentation of SMN, fibrillarin and the box C/D snoRNAs..... | 159 |
| 5.2.4. SMN is not a stable component of the U3 and U8 box C/D pre- snoRNPs..... | 161 |
| 5.2.5. Gemin5 is essential for the accumulation of box C/D snoRNA..... | 164 |
| 5.2.6. The SMN complex is essential for the correct localisation of box C/D snoRNA..... | 168 |
| 5.3. Discussion..... | 174 |

Chapter six: Analysis of the interaction of Snurportin1 with the box C/D pre- snoRNP.....179

| | |
|---|-----|
| 6.1. Introduction..... | 180 |
| 6.2. Results..... | 183 |
| 6.2.1. Snurportin1 interacts with the box C/D snoRNP common core protein NOP56..... | 184 |
| 6.2.2. Snurportin1 interacts with the box C/D snoRNP assembly factors TIP48 and TIP49..... | 185 |
| 6.2.3. Snurportin1 does not interact with the box C/D snoRNP assembly factors NUFIP, TAF9, NOP17 or the transport factor PHAX..... | 187 |
| 6.2.4. Snurportin1 does not interact with the SMN complex..... | 188 |
| 6.3. Discussion..... | 190 |

| | |
|--|------------|
| Chapter Seven: General discussion..... | 192 |
| 7.1. Overview..... | 193 |
| 7.2. Box C/D snoRNP biogenesis..... | 193 |
| 7.3. The SMN complex functions in the biogenesis of numerous RNPs..... | 196 |
| 7.4. Implications of study..... | 199 |
| 7.5. Future studies..... | 201 |
| References..... | 204 |
| Publications and presentations..... | 249 |

List of Figures

| | |
|--|----|
| Figure 1.1: The chemical modification of rRNA by the box C/D and box H/ACA SnoRNPs..... | 4 |
| Figure 1.2: The box C/D snoRNP..... | 7 |
| Figure 1.3: The eukaryotic box H/ACA snoRNP..... | 9 |
| Figure 1.4: The human U85 scaRNA..... | 11 |
| Figure 1.5: pre-rRNA processing in <i>S. cerevisiae</i> | 14 |
| Figure 1.6: Overview of ribosome biogenesis in humans and yeast..... | 15 |
| Figure 1.7: Modification sites on the yeast large ribosomal subunit..... | 16 |
| Figure 1.8: The human U3 box C/D snoRNP..... | 18 |
| Figure 1.9: Pseudoknot formation in the 18S rRNA of <i>S. cerevisiae</i> | 20 |
| Figure 1.10: Genomic organisation of human box C/D snoRNAs..... | 22 |
| Figure 1.11: Processing of an intronic box C/D snoRNA in yeast..... | 23 |
| Figure 1.12: A model of the 3' processing of the yeast U3 snoRNA..... | 25 |
| Figure 1.13: Schematic representation of the box C/D pre-snoRNP interaction network..... | 28 |
| Figure 1.14: Box C/D snoRNP biogenesis..... | 35 |
| Figure 1.15: Human U1 snRNP structure..... | 37 |

| | |
|--|-----|
| Figure 1.16: Pre-mRNA splicing..... | 39 |
| Figure 1.17: U1, U2, U4 and U5 snRNP biogenesis..... | 43 |
| Figure 1.18: SMN structure and localisation..... | 46 |
| Figure 1.19: Sub nuclear structures..... | 52 |
| Figure 1.20: The nucleolus..... | 53 |
| Figure 1.21: Cajal bodies..... | 54 |
| Figure 2.1: Tet-responsive U3 box C/D snoRNA promoter..... | 62 |
| Figure 2.2: Schematic representation of integrated pcDNA/FRT/TO plasmids..... | 66 |
| Figure 3.1: Nuclear localisation of the box C/D snoRNAs and fibrillarin..... | 84 |
| Figure 3.2: Box C/D snoRNP biogenesis..... | 85 |
| Figure 3.3: Depletion of fibrillarin..... | 88 |
| Figure 3.4: The localisation of box C/D snoRNA in cells depleted of the common core proteins..... | 89 |
| Figure 3.5: The localisation of box C/D snoRNA in cells depleted of hPRP31..... | 92 |
| Figure 3.6: The localisation of the box C/D snoRNP assembly factors..... | 94 |
| Figure 3.7: StreptoTag U3 snoRNA..... | 98 |
| Figure 3.8: The localisation of newly synthesised StreptoTag U3 snoRNA in NOP56 and NOP58 depleted cells..... | 100 |

| | |
|---|-----|
| Figure 3.9: Schematic diagram of the Tet-responsive StreptoTag U3 snoRNA..... | 102 |
| Figure 3.10: Development of a Tet-inducible mU14 snoRNA HEK293 cell line..... | 104 |
| Figure 3.11: Schematic diagram of the Tet-responsive fibrillarin and SMN HEK293 expression cassette..... | 105 |
| Figure 3.12: Tet-inducible FLAG-fibrillarin HEK293 cell line controls | 107 |
| Figure 3.13: Tet-inducible FLAG-SMN HEK293 cell line controls..... | 112 |
| Figure 3.14: Depletion of endogenous SMN in induced FLAG-SMN HEK293 cells. | 114 |
| Figure 3.15: Depletion of SMN in HeLa SS6 cells..... | 115 |
| Figure 3.16: Organisation of DD-fibrillarin plasmid..... | 116 |
| Figure 3.17: Control experiments DD-fibrillarin accumulation..... | 118 |
| Figure 3.18: Immunofluorescence of DD-fibrillarin..... | 120 |
| Figure 3.19: The localisation of DD-fibrillarin in NOP56 depleted cells..... | 122 |
| Figure 4.1: The chemical modification of rRNA by the box C/D snoRNPs..... | 131 |
| Figure 4.2: Protein sequence alignments of fibrillarin from various organisms..... | 132 |
| Figure 4.3: Crystal structure of archaeal fibrillarin..... | 133 |
| Figure 4.4: Differential nucleolar localisation of fibrillarin and BMS1 | 134 |
| Figure 4.5: Fibrillarin deletion mutants..... | 137 |
| Figure 4.6: The localisation of GFP-fibrillarin deletion mutants..... | 138 |

| | |
|---|-----|
| Figure 4.7: Interaction assay between fibrillarin and the box C/D snoRNP assembly factors NUFIP, BCD1 and NOP17..... | 141 |
| Figure 4.8: Interaction assay between fibrillarin and the box C/D snoRNP assembly factors TIP48 and TIP49..... | 142 |
| Figure 4.9: Interaction assay between the fibrillarin deletion mutants and BCD1, NOP17 and NUFIP..... | 144 |
| Figure 4.10: Interaction assay between the fibrillarin deletion mutants and the box C/D snoRNP assembly factors TIP48 and TIP49..... | 145 |
| Figure 5.1: The SMN complex..... | 150 |
| Figure 5.2: The role of the SMN complex in snRNP biogenesis..... | 152 |
| Figure 5.3: Differential nucleolar localisation of the box C/D snoRNAs..... | 153 |
| Figure 5.4: Interaction assay between fibrillarin and components of the SMN complex..... | 156 |
| Figure 5.5: Interaction assay between components of the SMN complex and the box C/D snoRNP assembly factors..... | 158 |
| Figure 5.6: Glycerol gradient analysis of HeLa SS6 nuclear extract..... | 160 |
| Figure 5.7: Immunoprecipitation of FLAG-SMN and FLAG-fibrillarin from HEK293 cell extracts..... | 162 |
| Figure 5.8: Depletion of SMN and Gemin2..... | 165 |
| Figure 5.9: Box C/D snoRNA levels after depletion of components of the SMN complex..... | 166 |

| | |
|---|-----|
| Figure 5.10: The localisation of box C/D snoRNA in SMN depleted cells..... | 169 |
| Figure 5.11: The localisation of box C/D snoRNAs in cells depleted of Gemin2, 5, 6 and 7..... | 171 |
| Figure 5.12: The localisation of 7SL in SMN depleted cells..... | 173 |
| Figure 6.1: The transport of snRNPs during biogenesis..... | 181 |
| Figure 6.2: Interaction assay between Snurportin1 and the box C/D snoRNP common core proteins..... | 184 |
| Figure 6.3: Interaction assay between Snurportin1 and the box C/D snoRNP assembly factors TIP48 and TIP49..... | 186 |
| Figure 6.4: Interaction assay between Snurportin1 and the box C/D snoRNP assembly factors and the transport factor PHAX..... | 188 |
| Figure 6.5: Interaction assay between Snurportin1 and components of the SMN complex..... | 189 |
| Figure 7.1: SMN could function as a transport factor in box C/D snoRNP biogenesis..... | 198 |

List of Tables

| | |
|--|----|
| Table 1.1: The box C/D snoRNP common core proteins..... | 6 |
| Table 1.2: The box H/ACA snoRNP common core proteins..... | 9 |
| Table 2.1: Primers used to create mU14 and StreptoTag U3 pc/DNA5/FRT/TO expression plasmids..... | 61 |
| Table 2.2: Primers used to create fibrillarin and SMN pcDNA/FRT/TO expression plasmids..... | 63 |
| Table 2.3: Primers used to create DD-fibrillarin reporter plasmid..... | 63 |
| Table 2.4: Primers used to create fibrillarin deletion mutants..... | 64 |
| Table 2.5: The siRNA duplexes..... | 68 |
| Table 2.6: Antibodies used in Western blot assays..... | 72 |
| Table 2.7: Primers used to create 7SL probe..... | 73 |
| Table 2.8: Oligonucleotides used in primer extension assays..... | 74 |
| Table 2.9: FISH probes..... | 76 |
| Table 2.10: Antibodies used in immunofluorescence..... | 77 |
| Table 2.11: Expression plasmids used to produce recombinant GST tagged proteins... | 78 |
| Table 2.12: Expression plasmids used to produce recombinant 6 x His tagged proteins..... | 79 |

| | |
|---|----|
| Table 2.13: Expression plasmids used to produce <i>in vitro</i> -translated ³⁵ S-labelled proteins..... | 80 |
|---|----|

Abbreviations

| | |
|-------------------|--|
| A | Alanine |
| AdoMet | S-adenosyl-L-methionine |
| AMV | Avian myeloblastosis virus |
| ATP | Adenosine triphosphate |
| BBP | Branch point binding protein |
| BCD1 | Box C/D snoRNA accumulation protein |
| BSA | Bovine serum albumin |
| C | Cytosine |
| C terminal | Carboxyl-terminus |
| <i>C. elegans</i> | <i>Caenorhabditis elegans</i> |
| CAB | Cajal Body box |
| CBC | Cap Binding complex |
| cDNA | Coding deoxyribonucleic acid |
| CMV | Cytomegalo virus |
| CRM1 | Chromosome maintenance region 1 |
| CTP | Guanosine triphosphate |
| D | Aspartic acid |
| DAPI | 4',6-diamidino-2-phenylindole |
| DD | degradation domain |
| DFC | Dense fibrillar centre |
| DMEM | Dulbecco's modified Eagle's medium |
| DNA | Deoxyribonucleic acid |
| DTT | Dithiothreitol |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | Ethylenediaminetetraacetic acid |
| ETS | External transcribed region |
| FBS | Foetal bovine serum |
| FC | Fibrillar centre |
| FISH | Fluorescent <i>in Situ</i> Hybridisation |
| FLAG-fibrillarin | 2 x FLAG 6 x His tagged fibrillarin |
| FLAG-SMN | 2 x FLAG 6 x His tagged SMN |

| | |
|----------------------|--|
| FRT | Flp Recombination Target |
| G | Glycine |
| G | Guanosine |
| GC | Granular component |
| GST | Glutathione S-Transferase) |
| GTP | Guanosine triphosphate |
| <i>H. sapiens</i> | <i>Homo sapiens</i> |
| HCL | Hydrochloric acid |
| HEPES | 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid |
| His | Histidine |
| His tag | 6 x Histidine tag |
| hnRNPs | Heterogenous nuclear ribonucleic protein complexes |
| HRP | Horse radish peroxidase |
| IGCs | Interchromatin granule clusters |
| IP | Immunoprecipitation |
| IPTG | Isopropyl-beta-D-thiogalactopyranoside |
| ITS | Internal transcribed region |
| K | Lysine |
| Kb | Kilobases |
| KCl | Potassium chloride |
| kDa | Kilodalton |
| KOAc | Potassium acetate |
| KOH | Potassium hydroxide |
| <i>M. jannaschii</i> | <i>Methanococcus jannaschii</i> |
| <i>M. thermo</i> | <i>Methanococcus thermophiles</i> |
| m ₃ G | 2,2,7-trimethylguanosine cap |
| m ⁷ G | 7-monomethylguanosine cap |
| MgCl ₂ | Magnesium Chloride |
| MgOAc | Magnesium acetate |
| MiRNAs | Micro ribonucleic acid |
| MOPS | 3-morpholinopropanesulfonic acid |
| MPP10 | M phase phosphoprotein 10 |
| mpppG | γ-monomethylphosphate |

| | |
|---------------------|--|
| mpppG | 5' γ -monomethylphosphate |
| mRNA | Messenger ribonucleic acid |
| N terminal | Amino terminus |
| NaCl | Sodium chloride |
| NaPO ₄ | Sodium phosphate |
| ncRNA | Non coding ribonucleic acid |
| NLS | Nuclear localisation signal |
| Nop17 | nucleolar localisation protein 17 |
| NORs | Nucleolar organiser regions |
| NPC | Nuclear pore complex |
| NUFIP | Nuclear FMRP interacting protein |
| P | Proline |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PHAX | Phosphorylated adaptor for RNA export |
| PML | Promyelocytic leukaemia |
| PNK | Polynucleotide kinase |
| pre-rRNA | Precursor ribosomal ribonucleic acid |
| pre-snoRNA | Precursor-small nucleolar ribonucleic acid |
| PRMT5 | Protein arginine methyltransferase5 |
| RISC | RNA-induced silencing complex |
| RNA | Ribonucleic acid |
| RNP | ribonucleic protein complexes |
| Rp | Ribosomal protein |
| rpm | Revolutions per minute |
| rRNA | Ribosomal ribonucleic acid |
| <i>S.cerevisiae</i> | <i>Saccharomyces Cerevisiae</i> |
| ScaRNPs | Small Cajal body RNPs |
| SDS | Sodium dodecyl sulphate |
| SIP1 | SMN interacting protein 1 |
| siRNA | Small interfering ribonucleic acid |
| SMA | Spinal muscular atrophy |

| | |
|------------------|---|
| SMN | Survival of motor neurons protein |
| SRP | Signal recognition particle |
| SSC | Sodium chloride- sodium citrate |
| SSU | Small subunit |
| T | Thymidine |
| TAF9 | Transcription activating factor 9 |
| TBP | TATA binding proyein |
| Tet | Tetracycline |
| TGS1 | Methyltransferase trimethylguanosine 1 |
| | TBP (TATA binding protein) - interacting protein of |
| TIP48 | 48kDa |
| | TBP (TATA binding protein) - interacting protein of |
| TIP49 | 49kDa |
| Tris | Tris(hydroxymethyl)aminomethane |
| tRNA | Transfer ribonucleic acid |
| TTP | Thymidine triphosphate |
| tUTP | Transcription required- U three protein |
| U2AF | U2 auxiliary factor |
| UNR | Upstream of N-ras |
| UNRIP | Unr (upstream of N-ras) -interacting protein |
| UTP | Uridine triphosphate |
| v/v | Volume per volume |
| w/v | Weight per volume |
| <i>X. Laevis</i> | <i>Xenopus laevis</i> |
| Xist | X-inactive-specific transcript |
| Y | Tyrosine |
| µg | Micogram |
| µl | Micolitre |
| µm | Micometer |

Chapter one

Introduction

1.1. Non-coding RNAs (ncRNAs)

Non-coding RNAs (ribonucleic acids) are functional RNAs that are not translated into proteins. The ncRNAs operate through base pairing to target sequences and have roles in a variety of cellular processes ranging from protein translation, pre-mRNA splicing, gene regulation and maintenance of chromosome structure.

A variety of ncRNAs are involved in protein translation and these include the ribosome, transfer RNAs (tRNAs), the signal recognition particle (SRP) and the small nucleolar RNAs (snoRNAs). The factory of protein translation is the ribosome, which catalyses the translation of nucleotide sequences into protein molecules. The eukaryotic ribosome consists of four ncRNAs known as ribosomal RNAs (rRNAs). The tRNAs function with the ribosome as adapters between amino acids and the messenger RNAs (mRNAs). The signal recognition particle (SRP), which contains an ncRNA component known as 7SL, is involved in the transportation of specific nascent proteins to the endoplasmic reticulum for modification. Another example of ncRNAs required for protein translation are the box C/D and H/ACA snoRNAs that function in the chemical modification and maturation of rRNA, which is required for the production of a functional ribosome.

In eukaryotes pre-mRNA splicing is performed by the spliceosome, which consists of ncRNAs known as small nuclear RNAs (snRNAs). The biogenesis of the snRNAs requires another type of ncRNA known as the small Cajal body-specific RNAs (scaRNAs) that function in the chemical modification of snRNA sequences.

The ncRNAs also have roles in gene regulation and maintenance of chromosome structure. Examples of ncRNAs that function in gene regulation are the micro RNAs (miRNAs) and the small interfering RNAs (siRNAs). These ncRNAs bind target mRNA sequences and prevent translation either through degradation of mRNA or blocking translation (Lui et al, 2005; Sen et al, 2005). It has been predicted that up to one third of all human protein synthesis is under miRNA control (Bentwich et al, 2005; Berezikov et al, 2005). Examples of ncRNAs that function in the maintenance of chromosome structure are the telomerase complex and Xist (X-inactive-specific transcript). The enzyme telomerase, which is responsible for telomere maintenance, contains an ncRNA

that functions as a template for elongating telomeres. The Xist ncRNA has roles in X chromosome inactivation through the recruitment of histone-modifying enzymes.

A lot of the ncRNAs discussed associate with proteins in what are termed ribonucleic protein complexes (RNPs). The ncRNAs and RNPs have been linked to a variety of human diseases, for instance the miRNAs and telomerase have been implicated in cancer and cardiovascular disease (Oulton and Harrington, 2000; Lu et al, 2005). The mutation of the U50 box C/D snoRNA has been linked to prostate cancer (Dong et al, 2008) while the snRNAs and spliceosome have been implicated in retinitis pigmentosa and spinal muscular atrophy (SMA; Vithana et al, 2001; Winkler et al 2005; Shpargel and Matera, 2005; Zhang et al, 2008). Furthermore, Diamond-Blackfan anemia (Flygare and Karlsson, 2006), Treacher Collins syndrome (Trainor et al, 2009) and a range of cancers (Dai and Lu, 2008) have been linked to ribosome biogenesis.

1.2. Small nucleolar ribonucleic protein complexes (snoRNPs)

The small nucleolar ribonucleic protein complexes (snoRNPs) function in the chemical modification of rRNA during ribosome biogenesis. These complexes consist of snoRNA and a number of common core proteins. In higher eukaryotes there are over 150 snoRNPs categorised into two main classes based on their function and structure: the box C/D and box H/ACA snoRNPs. It is, however, worthy of note that a few snoRNPs function as chaperones in precursor rRNA (pre-rRNA) processing (Omer et al, 2000).

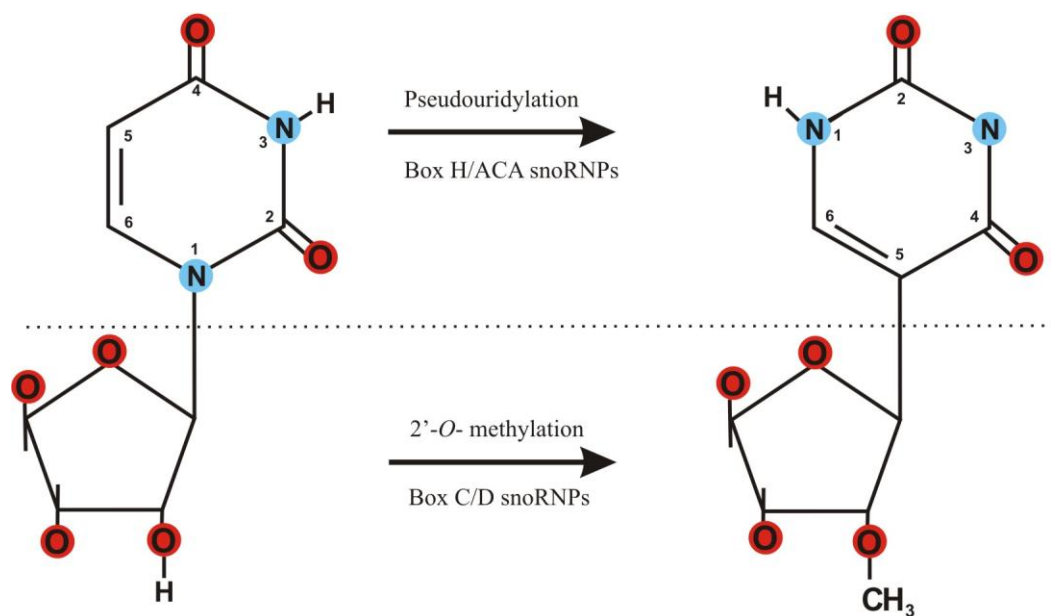


Figure 1.1: The chemical modification of rRNA by the box C/D and H/ACA snoRNPs

The top panel shows the conversion of uridine to pseudouridine performed by the box H/ACA snoRNPs. The bottom panel shows the 2'-O-methylation of the ribose sugar backbone of rRNA performed by the box C/D snoRNPs. Blue circles represent nitrogen while red circles correspond to oxygen. Black lines indicate carbon and covalent bonds.

Both box C/D and H/ACA snoRNPs have been identified in a range of eukaryotes from metazoans to plants, yeast, and kintoplastid protozoans. Interestingly, homologues of snoRNPs have also been found in archaeal, indicating that the snoRNPs are of an ancient origin

1.2.1. Box C/D snoRNPs

The eukaryotic box C/D snoRNPs direct the 2'-*O*-methylation of target rRNA residues and consist of box C/D snoRNA and the common core proteins: fibrillarin, NOP56, NOP58 and 15.5K (Tyc and Steitz, 1989; Lyman et al, 1999; Lafontaine & Tollervey, 1999, 2000; Newman et al, 2000; Watkins et al, 2000; Figure 1.2 and Table 1.1).

The box C/D snoRNAs are characterised by the conserved C box (RUGAUGA, where R is purine) and D box (CUGA), which are located near the 5' and 3' end of the snoRNA, respectively (Figure 1.2). In most human box C/D snoRNAs the C and D boxes are brought together through base pairing of the terminal 5' and 3' nucleotides to form a terminal stem loop structure. Many yeast (and some human) box C/D snoRNAs, however, lack a terminal stem loop and the C/D box is formed through transient external or stable internal base-pairing (Darzacq and Kiss, 2000). The majority of box C/D snoRNPs also contain another set of less conserved motifs, the C' and D' boxes, located in the central region of the snoRNA (Kiss-Laszlo et al, 1998). The box C/D snoRNAs contain a guide sequence located adjacent to the D and / or D' box which binds target rRNA sequences. Once bound to the snoRNA, the rRNA nucleotide base-paired five nucleotides upstream of the conserved CUGA motif is methylated (Kiss-Laszlo et al, 1996; 1998; Figure 1.2).

Fibrillarin is the methyltransferase component of the box C/D snoRNPs and transfers a methyl group from the co-factor S-adenosyl-L-methionine (AdoMet) to the donor site located on the ribose sugar of the target rRNA residue (Tollervey et al, 1993; Figure 1.1). The related common core proteins NOP56 and NOP58 have 45% sequence identity and belong to a family of proteins, which includes the U4 small nuclear RNP (snRNP) core protein hPRP31, that contain a conserved NOP domain proposed to function in RNP association (Gautier et al, 1997). Both NOP56 and NOP58 are essential for ribosome biogenesis; however, their exact function is not clear (Gautier et al, 1997). The 15.5K common core protein belongs to the L7Ae protein family that share a homologous RNA binding domain, which includes the box H/ACA snoRNP core protein NHP2 (Watkins et al, 1998) and a number of ribosomal proteins (Koonin et al, 1994). The association of 15.5K to the box C/D snoRNA is essential for the binding of

the other common core proteins and it has been proposed that the binding of 15.5K results in a conformational change in the snoRNA providing additional binding sites (Watkins et al, 2002).

Much of the box C/D snoRNP structural data has been produced in studies of archaeal box C/D snoRNPs, which contain homologues of fibrillarin, 15.5K (L7Ae in archaea) and a single homologue of NOP56 and NOP58 known as Nop5 (Table 1.1). Interestingly, L7Ae is also a core protein of the archaeal box H/ACA snoRNPs and a ribosomal protein (Khun et al, 2002). Analysis of the archaea proteins suggest that the box C/D snoRNPs adopt a symmetrical structure with a copy of fibrillarin, Nop5 and L7Ae located at both the C/D and C'/D' box (Aittaleb et al, 2003; Rashid et al, 2003; Tran et al, 2003; Figure 1.2 A). This model was indicated, in part, by the crystallisation of a tetrameric complex consisting of archaeal Nop5 and fibrillarin. In this complex two dimers of fibrillarin and Nop5 associate with one another through the coiled-coil domains of Nop5 (Aittaleb et al, 2003). This model provides a mechanism by which fibrillarin can associate at both catalytic sites of the snoRNP and direct two independent modifications (Figure 1.2 A). Interestingly, the coiled-coil domains of Nop5 are not required for the assembly of the archaeal box C/D snoRNP; however, these domains are essential for the function of the complex (Zhang et al, 2006). This indicates that the self-dimerisation of Nop5p within the box C/D snoRNP may be important for structure positioning fibrillarin.

| Humans | Yeast | Archaea |
|-------------|--------|-------------|
| fibrillarin | Nop1p | fibrillarin |
| NOP56 | Nop56p | Nop5 |
| NOP58 | Nop58p | |
| 15.5K | Snu13p | L7Ae |

Table 1.1: The box C/D snoRNP common core proteins

The table shows the box C/D snoRNP common core protein homologues from humans, yeast and archaea. Archaea contain a single homologue of the eukaryotic NOP56 and NOP58 proteins known as Nop5.

The eukaryotic box C/D snoRNPs are proposed to adopt a pseudo-symmetrical model similar to the archaeal box C/D snoRNPs. The *in vitro* construction of the *Xenopus* U25 box C/D snoRNP, followed by UV cross-linking, indicates that 15.5K and NOP58 associate with the C/D box while NOP56 binds the C'/D' box (Figure 1.2 B; Cahill et al, 2002). As with the archaeal box C/D snoRNPs, fibrillarin was shown to associate with the *Xenopus* U25 box C/D snoRNP at both the C/D and C'/D' boxes, indicating two copies of fibrillarin are present (Figure 1.2 B). The positioning of fibrillarin at both the C/D and C'/D' boxes is also in agreement with functional data that some box C/D snoRNPs can perform two independent rRNA modifications (Kiss-Laszlo et al, 1998; Figure 1.2 B).

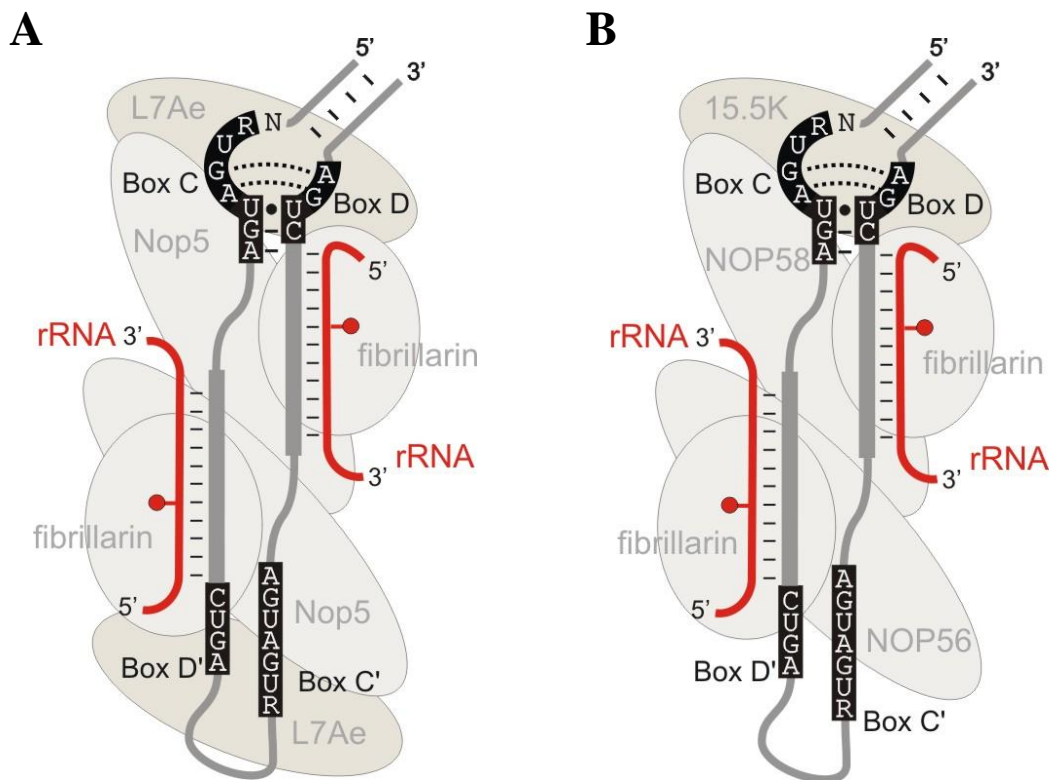


Figure 1.2: The box C/D snoRNP

Schematic diagram of the structure of (A) an archaeal (B) an eukaryotic box C/D snoRNP. The grey line represents the box C/D snoRNA with the conserved sequences shown in white on black boxes. The thicker grey lines represent the guide region. Red lines represent the pre-rRNA with the methylation site indicated by a sphere. The core proteins are shown as grey ovals behind the snoRNA. Image provided by N.J.Watkins, Newcastle University, UK.

1.2.2. Box H/ACA snoRNPs

The eukaryotic box H/ACA snoRNPs direct pseudouridylation of target rRNA residues (Figure 1.1) and consist of snoRNA and the common core proteins: NHP2, NOP10, GAR1 and dyskerin (Watkins et al, 1998; Henras et al, 1998). Dyskerin is responsible for the pseudouridylation of target rRNA residues (Lafontaine et al, 1998; Zebajarian et al, 1999).

The box H/ACA snoRNAs are characterised by the conserved H (ANANNA, with N being any nucleotide) and ACA boxes. The majority of box H/ACA snoRNAs form a twin hairpin structure with the conserved H box located in a hinge region separating the hairpins (Figure 1.3). A minority of box H/ACA snoRNAs form a single hairpin structure. Whether the box H/ACA snoRNA forms a single or double hairpin structure the conserved ACA box is located three nucleotides upstream from the 3' end. Within the snoRNA hairpins there are internal RNA loops, commonly known as pseudouridylation pockets, which contain between 9 and 13 nucleotides on each strand that bind target rRNA (Figure 1.3). Once the target rRNA is bound pseudouridylation occurs around 14-16 bases upstream of either the H or ACA box depending upon the specific snoRNP (Ganot et al, 1997; Ni et al, 1997; Balakin et al, 1996).

The majority of the structural data on box H/ACA snoRNPs comes from studies of archaea; however, eukaryotic box H/ACA snoRNPs are predicted to have a similar structure (Watkins et al, 1998). The archaeal box H/ACA snoRNPs contain homologues of all the human core proteins; however, in archaea dyskerin is known as Cbf5 and NHP2 is referred to as L7Ae (Table 1.2). Analysis of the crystal structure of a single hairpin archaeal box H/ACA snoRNP revealed that the upper loop of the snoRNA hairpin binds a composite surface made up of L7Ae, Nop10 and Cbf5 (Li and Ye, 2006; Figure 1.3). Cbf5 was also shown to bind to the ACA box, which with the interaction with the upper RNA loop positions the catalytic domain of Cbf5 across the pseudouridylation pocket (Li and Ye, 2006). Archaeal Gar1 is the only box H/ACA core protein that does not bind directly to the snoRNA and crystallisation studies as well as biochemical studies indicate Gar1 interacts with directly with Cbf5, where it is presumed to function in substrate loading and release (Baker et al, 2005, Charpentier et

al, 2005; Henras et al, 2004; Rashid et al, 2006; Hamma et al, 2005; Manival et al, 2006; Li and Ye, 2006).

| Humans | Yeast | Archaea |
|----------|--------|---------|
| dyskerin | Cbf5p | Cbf5 |
| GAR1 | Gar1p | Gar1 |
| NOP10 | Nop10p | Nop10 |
| NHP2 | Nhp2p | L7Ae |

Table 1.2: The box H/ACA snoRNP common core proteins

The table shows the box H/ACA snoRNP common core protein homologues from humans, yeast and archaea.

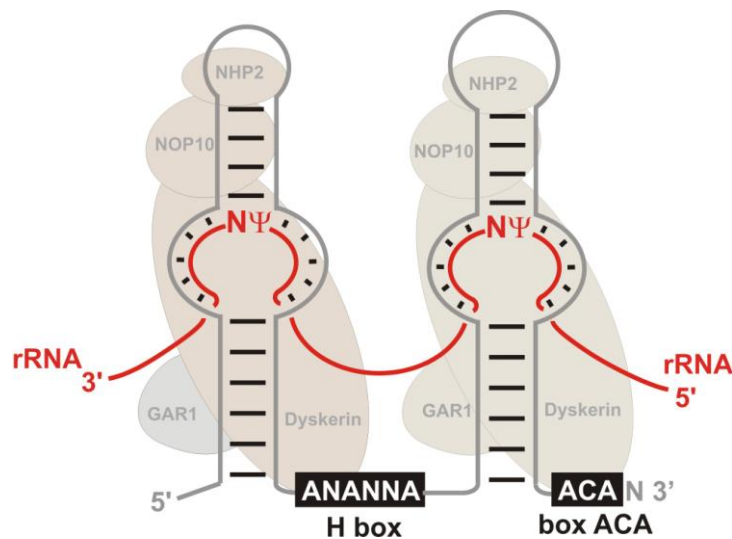


Figure 1.3: The eukaryotic box H/ACA snoRNP

The grey line represents the box H/ACA snoRNA with the conserved sequences shown in white on black boxes. The thicker grey lines represent the guide region. A red line represents the pre-rRNA with the pseudouridylation site indicated by N ψ . The core proteins are shown as grey spheres behind the snoRNA. Figure adapted from Reichow et al, 2007.

Box H/ACA snoRNAs containing a twin hairpin structure have been predicted to associate with two sets of the core proteins (Watkins et al, 1998; Figure 1.3). Interestingly, mutation of the single ACA box of box H/ACA snoRNAs with a twin hairpin structure results in the loss of both copies of Cbf5, indicating that common core protein association is coupled (Baker et al, 2005).

1.2.3. The small Cajal body-specific RNPs (scaRNPs)

The small Cajal body-specific RNPs (scaRNPs) function in the chemical modification of spliceosomal snRNAs. The scaRNPs are functionally and structurally similar to the snoRNPs and perform 2'-*O*-methylation and pseudouridylation of target snRNA residues (Kiss, 2001; Kiss et al, 2002; Darzacq et al, 2002). The chemical modifications performed by the scaRNPs are mostly found in regions of the snRNAs involved in the spliceosomal RNA-RNA interactions and are vital for their function (Zhao and Yu, 2004).

The scaRNPs are composed of scaRNA associated with snoRNP common core proteins. The scaRNAs that direct 2'-*O*-methylation contain the conserved C (RUGAUGA, where R is purine) and the D (CUGA) boxes, as well as the less conserved internal C' and D' boxes. The scaRNAs that guide pseudouridylation contain the conserved H (ANANNA, with N being any nucleotide) and ACA boxes (Kiss, 2001). The box C/D and H/ACA scaRNAs have been predicted to associate with the respective core proteins of the box C/D and H/ACA snoRNPs, although to date only the association of fibrillarin and GAR1 has been demonstrated experimentally (Darzacq et al, 2002). Interestingly, a number of scaRNAs have been shown to contain both the C/D and H/ACA box motifs, for example the U85 scaRNP (Figure 1.4), and these scaRNPs can direct both the 2'-*O*-methylation and pseudouridylation of target snRNA residues (Jady and Kiss, 2001). It is proposed that the scaRNPs, which contain both the C/D and H/ACA box motifs, associate with both sets of snoRNP core proteins (Jady and Kiss, 2001).

The scaRNPs function in the modification of snRNA specifically in the Cajal bodies and their localisation is regulated by a Cajal body retention element termed the CAB box (Cajal Body box). The sequence of the CAB box is UGAG with the first and second bases varying amongst scaRNPs (Richard et al, 2003).

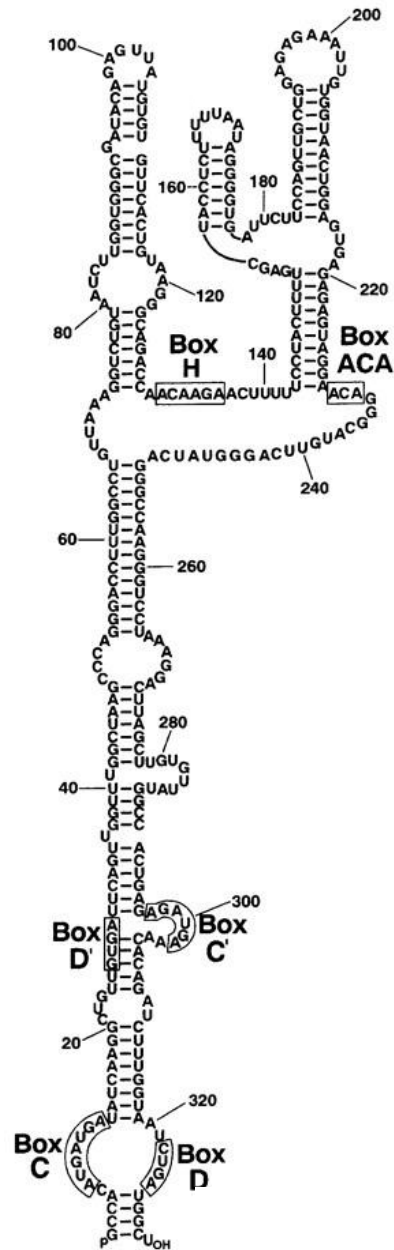


Figure 1.4: The human U85 scaRNA

Computer predicted two-dimensional structure of the human U85 scaRNA. The scaRNA nucleotide sequences are shown in black lettering. The positions of the conserved C, D, H and ACA boxes are labelled and indicated by black outlined boxes. Numerical values correspond to nucleotide number. Image taken from Richard et al, 2003.

1.3. Ribosome biogenesis

The eukaryotic ribosome is composed of two subunits, the 40S (small ribosomal subunit) and the 60S (large ribosomal subunit), each consisting of different rRNAs and ribosomal proteins. The 40S subunit contains a single rRNA species, known as 18S, while the 60S subunit contains three rRNA species, referred to as 5S, 5.8S and 25S / 28S (yeast / human). In eukaryotes the 5S rRNA is synthesised as a single transcript in the nucleoplasm. However, the 18S, 5.8S and 25 / 28S rRNAs are transcribed in the nucleolus as a large single pre-rRNA known as the 35S in yeast and the 47S in humans (Lafontaine and Tollervey, 2001). As the large pre-rRNA transcript migrates through the nucleolus it is chemically modified (Figure 1.1) and processed to produce the mature length rRNAs.

1.3.1. rDNA gene organisation and transcription

In humans there are between 300 and 400 copies of the 47S rDNA gene, which are transcribed by RNA polymerase I. The 47S rDNA genes are arranged in tandem repeats of around 13 kilobases (Kb) in size and are found on the p arms of chromosomes 13, 14, 15, 21 and 22 (Grummt and Pikaard, 2003; Russell & Zomerdijs, 2005). In the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) there are around 150 copies of the 35S rDNA gene arranged in tandem repeats on chromosome 12 and transcribed by RNA polymerase I. The clusters of rDNA genes are known as nucleolar organiser regions (NORs) and recruit factors required for transcription and processing of the rRNA resulting in the formation of the nucleolus. Several rDNA clusters from different chromosomes often gather together to form a single nucleolus (Schwarzacher and Mosgoeller, 2000).

The nucleolus consists of three sub-compartments, the inner fibrillar centre (FC), the dense fibrillar component (DFC) and the granular component (GC) (Schwarzacher and Mosgoeller, 2000). The rDNA genes are located, either within, or at the periphery of the FC and are transcribed at the border between the FC and DFC (Chooi and Leiby, 1981). As the pre-rRNA transcript migrates through the DFC and then the GC it is modified and processed by factors that are localised to these compartments (Figure 1.5 and 1.6).

The human 5S rDNA genes are found in tandem repeats separate from the 47S rDNA genes on the q arm of chromosome 1 and are transcribed by RNA polymerase III in the nucleoplasm (Bogdanov et al, 1995). In *S. cerevisiae* the 5S genes are located between each tandem repeat of the 35S rDNA, but are independently transcribed by RNA polymerase III (Lafontaine and Tollervey, 2001; Granneman and Baserga, 2004).

1.3.2. Processing of the large pre-rRNA transcript

Due to the ease of manipulation and genetic suitability the most extensive study of pre-rRNA processing has been performed using the yeast *S. cerevisiae* (Venema and Tollervey, 1999; Granneman and Baserga, 2004; Figure 1.5); however, pre-rRNA processing in higher vertebrates is similar. The processing of pre-rRNA is performed by a variety of endo and exonucleases and a few snoRNPs, which function as chaperones (Granneman and Baserga, 2004).

In *S. cerevisiae* the 18S, 5.8S and 25S rRNAs are transcribed as long 35S pre-rRNAs that contain 5' and 3' external transcribed regions (ETS) as well as two internal transcribed regions (ITS1 and 2; Figure 1.5). Pre-rRNA processing begins with initial cleavages at the A₀ (5' ETS) and the A₁ sites (5' ETS / 18S boundary) resulting in the removal of the 5' ETS sequence and forming the 32S precursor. Cleavage at the A₂ site (ITS1) produces two precursors, the 20S precursor that contains 18S, and the 27SA₂ precursor that contains 25S and 5.8S (Figure 1.5). The 20S precursor is transported to the cytoplasm where cleavage at the D site produces the mature 18S rRNA (Fatica et al, 2003; Schafer et al, 2003). Two alternative pathways process the 27SA₂ precursor resulting in the formation of two slightly different 5.8 rRNAs (Figure 1.5). Around 85 % of the 27SA₂ is cleaved at the A₃ site (ITS1) and then trimmed by exonucleases to the B_{1S} site. The other 15 % (of 27SA₂) is cleaved at B_{1L} site. After this point processing in both pathways is identical with cleavage at the B₂ site (3' ETS) and B₁ site (Figure 1.5). Due to the alternative cleavages earlier on two 27SB precursors are produced, designated S or L. Both the 27S transcripts are processed from C₂ to C₁ and are also subjected to 3'-5' exonucleolytic digestion of the ITS2 to site E. The end result of the

two pathways is the production of identical 25S rRNA and two forms of 5.8S, known as 5.8S_S and 5.8S_L.

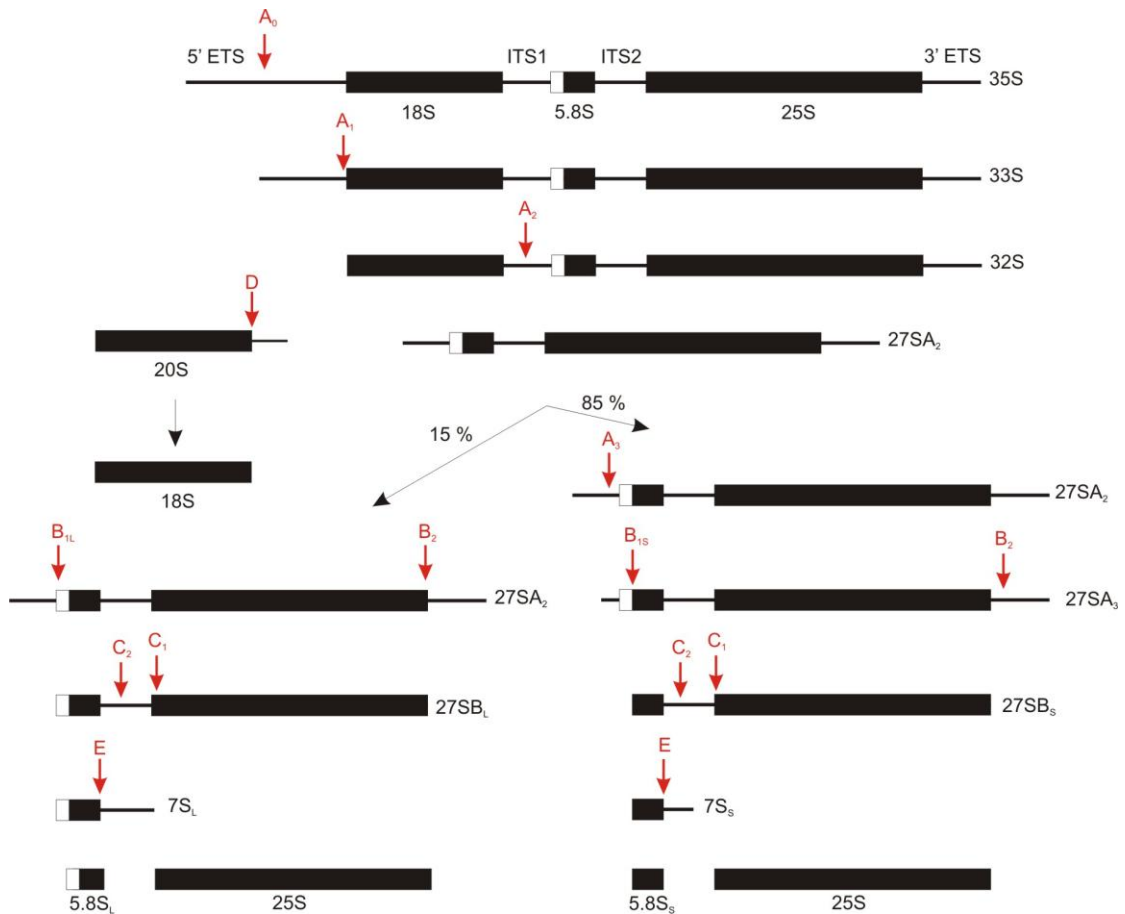


Figure 1.5: pre-rRNA processing in *S. cerevisiae*

The *S. cerevisiae* 18S, 5.8S and 25S rRNA are transcribed as a single 35S pre-rRNA transcript which is shown at the top of the panel. Black lines represent external and internal transcribed regions (ETS and ITS, respectively), which are removed during processing. Solid black boxes represent the 18S, 5.8S and 25S rRNAs (as indicated). Red arrows and red lettering indicate cleavage and processing sites. Black arrows represent progression of processing. Figure adapted from Granneman and Baserga, 2004.

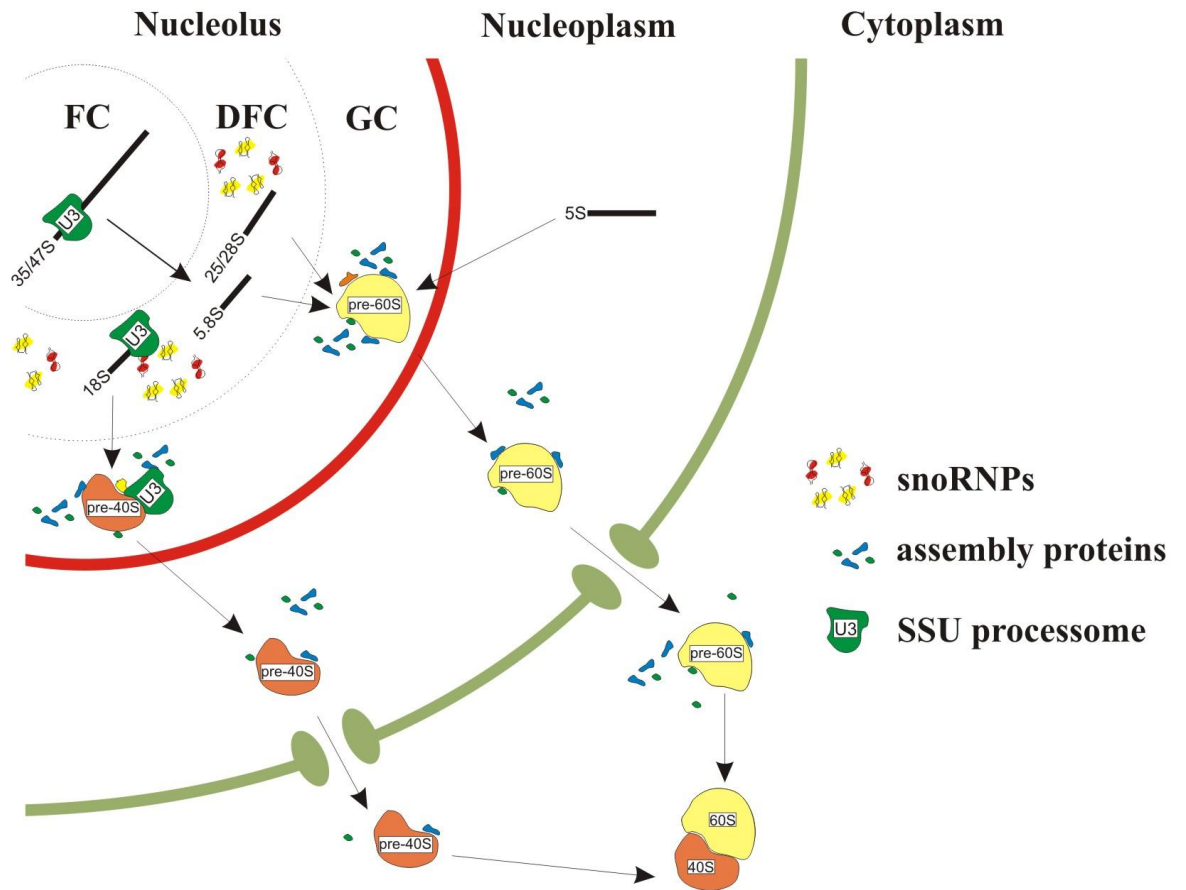


Figure 1.6: Overview of ribosome biogenesis in humans and yeast

The rRNA is represented as a black line with a label indicating its species. A red line represents the nucleolar boundary. The FC (fibrillar centre), DFC (dense fibrillar centre) and GC (granular component) are separated by black dotted lines. A green line represents the nuclear membrane. Green ovals correspond to nuclear pores. Black arrows show progression of biogenesis. Ribosomal subunits, snoRNPs and SSU processome shown as coloured shapes.

1.3.3. Chemical modification of rRNA

The most common chemical modifications on rRNA are the methylation of the 2'-hydroxyl of ribose sugars (2'-*O*-methylation) and the isomerization of uridines to pseudouridines (pseudouridylation) (Figure 1.1), which are performed by the box C/D and H/ACA snoRNPs, respectively. These chemical modifications occur in the DFC of the nucleolus in tandem with the cleavage / processing of the large pre-rRNA (Figure 1.6).

The number of rRNA modifications performed by the snoRNPs varies in different organisms, for example the rRNA of *S. cerevisiae* has 54 methylation and 44 pseudouridylation sites, while human rRNA has 105 methylation and 91 pseudouridylation sites (Bakin, 1998; Ofengand and Bakin, 1997; Maden, 1990).

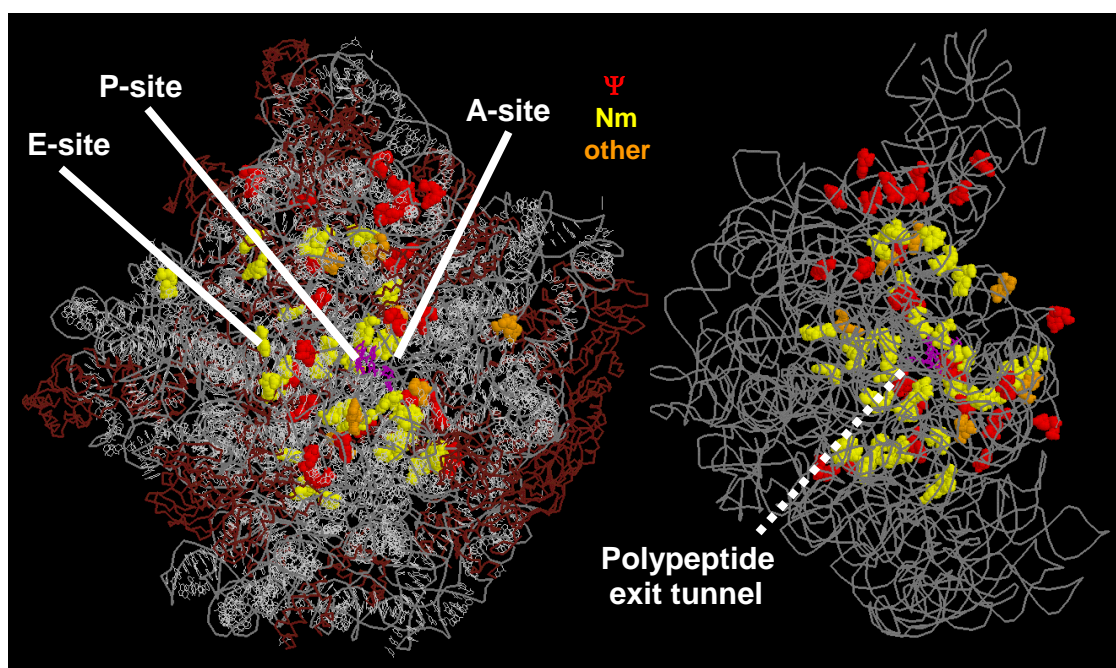


Figure 1.7: Modification sites on the yeast large ribosomal subunit

Two views of the yeast large ribosomal subunit showing pseudouridylation sites (Ψ , red), 2'-*O*-methylation sites (Nm, yellow) and other modifications (orange). P, E and A sites of tRNA and mRNA binding and the polypeptide exit tunnel are indicated. Modifications are shown by displaying the full van der Waals radii, rRNA backbone displayed in grey and protein chains maroon. Image provided by Decatur and Fournier, 2002.

The three-dimensional representation of the locations of modified rRNA nucleotides in the yeast ribosome revealed that they are in areas of functional importance. These include the peptidyl transferase centre, the P, E and A sites of tRNA and mRNA binding, the polypeptide exit tunnel and sites of ribosome subunit-subunit interactions (Figure 1.7; Decatur and Fournier, 2002). The positions of the chemical modifications are highly conserved and have been proposed to function as “fine tuners” of rRNA structure (Ofengand et al, 2002; Decatur and Fournier, 2002; King et al, 2003). The global inhibition of rRNA modifications results in growth defects in yeast indicating that rRNA modification is essential for the biogenesis and function of the ribosome (Tollervey et al, 1993; Zebajadian et al, 1999; King et al, 2003). It has been revealed that a few modification sites are of more functional importance than others and the inhibition of modification on these sites results in growth defects in yeast (Esguerra et al, 2008). However, more precise methods of analysis were required to determine the effect of individual chemical modifications.

1.3.4. A number of snoRNPs function as chaperones in pre-rRNA processing

A number of the box C/D and H/ACA snoRNPs are essential for the processing of the large pre-rRNA in both yeast and higher eukaryotes. These snoRNPs are predicted to function by base pairing to specific pre-rRNA regions and orchestrating the folding of the pre-rRNA into a conformation that allows efficient processing. The U3 (Hughes and Ares, 1991), U14 (Li et al, 1990; Liang and Fournier, 1995) and U22 box C/D snoRNPs (Tykocwski et al, 1994), as well as the snR10 (Tollervey, 1987), snR30 (U17 / E2; Morrissey and Tollervey, 1997) and E3 (Mishra and Eliceiri, 1997) box H/ACA snoRNPs are all required for the production of the 18S rRNA. The vertebrate specific U8 box C/D snoRNP is essential for the processing of 5.8S and 28S and is also predicted to be involved in the structural rearrangements required for the production of the two mature rRNAs (Peculis and Steitz, 1993).

The most studied of the snoRNPs that function in pre-rRNA processing is the U3 box C/D snoRNP. The U3 snoRNA contains two functional regions designated the 5' and the 3' domains. The 5' domain consists of sequence elements that bind to target pre-rRNA sequences, such as the GAC box, A box, A' box, 5' hinge and 3' hinge (Terns

and Terns, 2002; Figure 1.8). The 3' domain of the U3 snoRNA contains the sequence elements to which the core proteins associate. Interestingly, the U3 snoRNA contains two unique elements one being the C'/D box, which associates with the core proteins 15.5K, NOP56, NOP58 and fibrillarin, and the other being the B/C box, which binds a second copy of 15.5K and the U3 specific hU3-55K protein (Watkins et al, 2000; Grannemen et al, 2002; 2004).

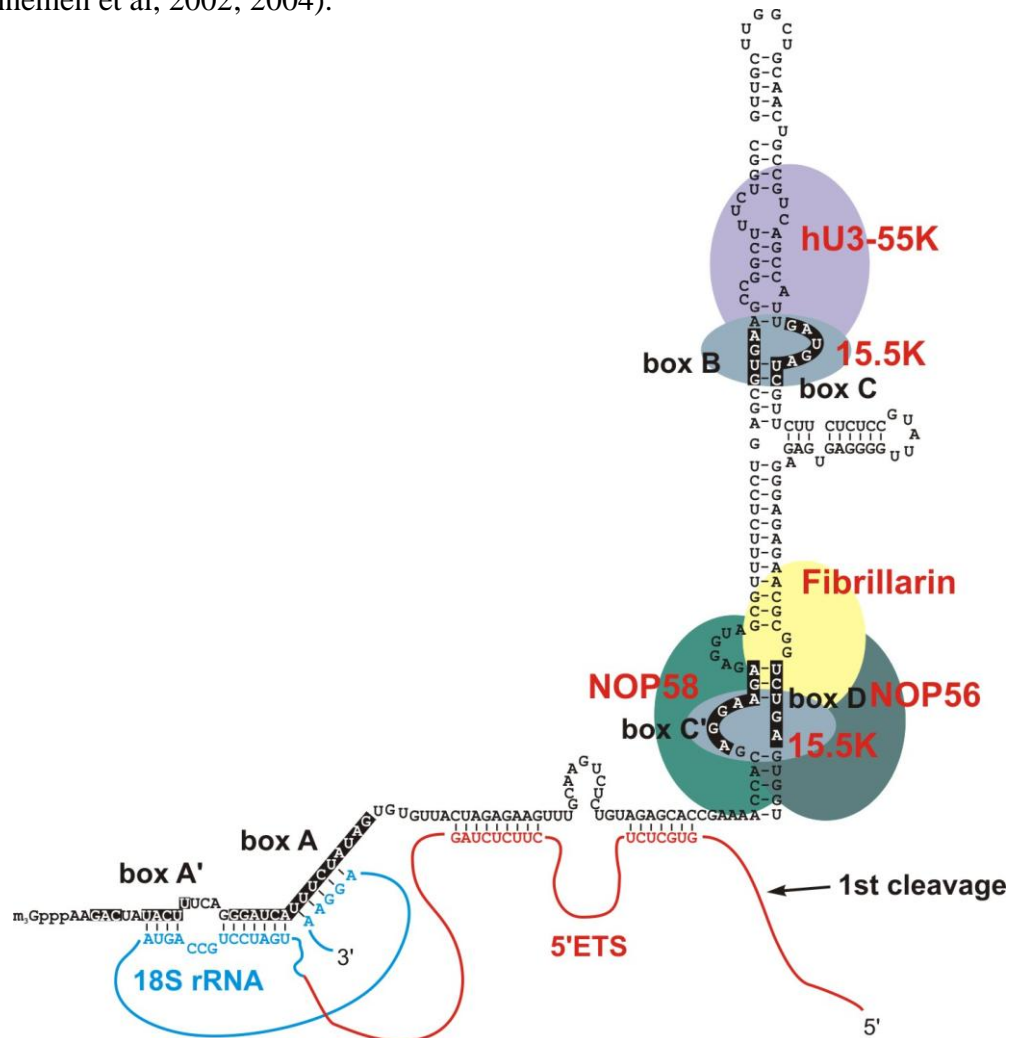


Figure 1.8: The Human U3 box C/D snoRNP

The U3 snoRNA is shown in black lettering except for conserved box motifs that are shown as white lettering on black background. The U3 snoRNA 5' hinge region base pairs with 5'ETS (red) of pre-rRNA while the A and A' boxes base pair with the 18S rRNA (blue). The 3' domain of U3 snoRNA contains the C'/D box, to which 15.5K (grey oval), NOP56 (dark green oval), NOP58 (light green oval) and fibrillarin (yellow oval) associate, and the U3 specific B/C box, which binds a second copy of 15.5k and the U3 specific hU3-55K protein (purple oval). The arrow shows first cleavage site of pre-rRNA. Image provided by N.J.Watkins, Newcastle University, UK.

Analysis of the interactions of yeast U3 snoRNA with the 35S pre-rRNA revealed that the interaction of the U3 snoRNA A box with the 5'ETS of the pre-rRNA is essential for cleavages at A₀, A₁ and A₂ sites while the interaction of the U3 snoRNA hinge region with 18S is required for cleavages at sites A₁ and A₂ (Beltrame et al, 1994; Beltrame & Tollervey, 1995; Sharma & Tollervey, 1999; Figure 1.8). The U3 snoRNP does not directly cleave the pre-rRNA but guides the activities of a complex known as the small subunit (SSU) processome.

The yeast SSU processome forms co-transcriptionally on the 35S pre-rRNA forming large 90S complexes, also known as 5' terminal knobs, which are visible in miller chromatin spreads of actively transcribing rRNAs (Miller & Beatty, 1969; Dragon et al, 2002). The SSU processome is a dynamic complex, which in yeast, has been shown to contain over 40 proteins including RNA helicases, RNA annealing factors and GTPases (Dragon et al, 2002). The five main sub-complexes that make up the SSU processome are the U3 snoRNP, MPP10 (M phase phosphoprotein 10) complex, tUTP (transcription required- U three protein) complex, bUTP complex and cUTP complex (Dragon et al, 2002). It has been proposed that the SSU processome functions in cleavages at the A₀, A₁ and A₂ pre-rRNA sites through the recruitment of endonucleases (Granneman et al, 2004). Although no endonucleases have been found associated with the SSU processome these factors may transiently bind and therefore be difficult to detect. Another possibility is that the interaction of the SSU processome with the pre-rRNA forms a ribozyme structure that directly cleaves the pre-rRNA (Collins and Guthrie, 2000). In addition to the cleavages that produce 18S rRNA the SSU processome also associates with factors required for the synthesis of the 40S ribosomal subunit. For instance the SSU processome contains the 40S ribosomal subunit proteins: Rps4 (ribosomal protein s4), Rps6, Rps7, Rps9 and Rps14, which become permanently associated with the 18S rRNA (Bernstein et al, 2004).

The U3 box C/D snoRNP also functions in orchestrating the folding of the 18S rRNA into its mature conformation. The U3 snoRNP base pairs to the 5' terminus of the 18S rRNA, via its A' and A boxes, and brings the 18S sequences into close proximity resulting in the formation of the conserved pseudoknot structure (Hughes, 1996; Figure 1.9).

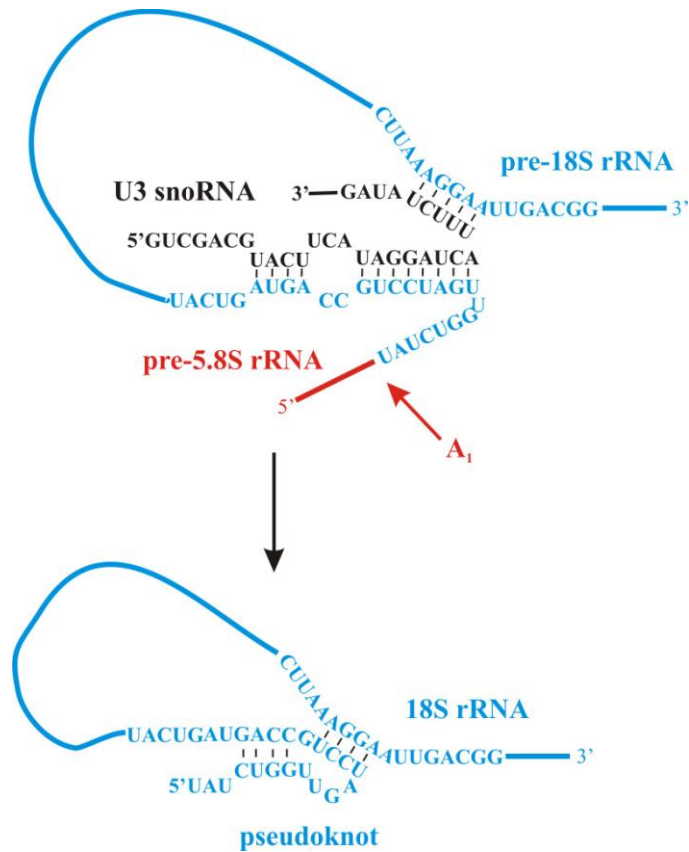


Figure 1.9: Pseudoknot formation in the 18S rRNA of *S. cerevisiae*

Interaction between the 5' domain of the yeast U3 box C/D snoRNA (black lettering) and the 5' end of yeast 18S rRNA (blue lettering and lines) still associated in 35S pre-rRNA. The 5'ETS of rRNA is shown (red lines) with A₁ cleavage site (located at 5'ETS and 18S boundary) represented as a red arrow. Short black lines represent bonds. The black arrow shows progression of process. Adapted from Hughes, 1996.

1.3.5. Ribosomal assembly and export

The mature 5.8S and 25S / 28S rRNAs (yeast / human) associate with the 5S rRNA in the GC of the nucleolus to form the large pre-60S complex, while the 18S rRNA forms the small pre-40S complex (Figure 1.6). The pre-60S and pre-40S ribosomal complexes are transported through the nucleoplasm to the cytoplasm during which further maturation events occur and ribosomal proteins associate (Figure 1.6; Tschöchner and Hurt, 2003). Once the final maturation steps have been performed the large 60S and smaller 40S ribosome subunits interact with an mRNA transcript to form a functional ribosome (Figure 1.6).

1.4. Box C/D snoRNP biogenesis

Box C/D snoRNP biogenesis proceeds in a large, dynamic multiprotein complex known as the pre-snoRNP that contains snoRNA, common core proteins and assembly, transport and RNA processing factors (Boulon et al, 2004; Watkins et al, 2004; 2007; McKeegan et al, 2007). The cellular localisation(s) where box C/D snoRNP biogenesis occurs is not clear with evidence suggesting the nucleoplasm (Terns et al, 1995; Watkins et al, 2004; Boulon et al, 2004), Cajal bodies (Verheggen et al, 2001; Boulon et al, 2004; Lemm et al, 2006) and cytoplasm (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007).

1.4.1. Gene organisation and transcription

Box C/D snoRNAs have a diverse gene organisation that differs between organisms. In vertebrates the majority of box C/D snoRNAs are encoded within the introns of housekeeping genes and are released by the processing of the excised introns from pre-mRNA (Kiss and Filipowicz, 1995). In yeast the majority of box C/D snoRNA genes are encoded as independent genes, being both monocistronic and polycistronic.

Two examples of human box C/D snoRNAs that have different genomic organisations are the U3 and U14 snoRNAs (Figure 1.10). In the human genome there are four to six copies of the U3 snoRNA, which are encoded as monocistronic genes with independent promoters. The U3 promoter region consists of a proximal sequence element (PSE) (instead of a TATA box), located 55 bases upstream of the U3 snoRNA, and a distal sequence element (DSE), around 220 bases upstream of the snoRNA sequence (Figure 1.10). The PSE site determines the transcription start site while the DSE functions as an enhancer (Mazan et al, 1993). The U3 snoRNA is transcribed by RNA polymerase II. The U14 snoRNA is encoded within introns 5, 6 and 7 of the heat shock cognate protein 70 (hsc70; Figure 1.10) and there are also two copies within the introns of ribosomal protein genes, which are all transcribed by RNA polymerase II (Lui and Maxwell, 1990; Kenmochi et al, 1996). As the U14 snoRNA is an intronic RNA its expression is based upon the expression of the host gene. Once transcribed the intronic box C/D snoRNAs are excised from the pre-mRNA.

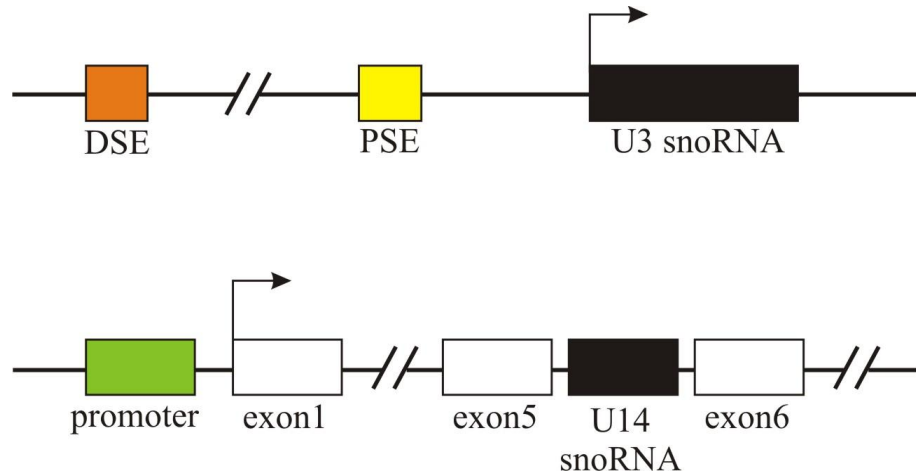


Figure 1.10: Genomic organisation of human box C/D snoRNAs

The top panel shows the organisation of the independently transcribed U3 box C/D snoRNA. The DSE is represented as an orange box and the PSE as a yellow box. U3 snoRNA is shown as a black box. The bottom panel shows the gene organisation of the intronic U14 snoRNA, which is expressed from the introns of the *hsc70* gene. The promoter is shown as a green box, the exons as white boxes and the U14 snoRNA as a black box. Arrows represent transcription start sites.

1.4.2. Processing of box C/D snoRNAs

Excision of intronic box C/D snoRNAs from host pre-mRNA occurs by one of two pathways, the splicing dependent and the splicing independent. The majority of intron-encoded snoRNAs are excised using the splicing dependent pathway where the intron is released from the pre-mRNA as a normal spliced lariat and then processed (Kiss and Filipowicz, 1995; Figure 1.11). In the splicing independent pathway, which was originally shown for the *Xenopus laevis* (*X. laevis*) U16 box C/D snoRNA, the snoRNA is released from unspliced pre-mRNA by endonucleolytic cleavages of flanking intronic sequences (Caffarelli et al, 1996). The snoRNA produced from both intronic and independent box C/D snoRNAs contain extended precursor sequences (pre-snoRNA) that are processed to form mature length box C/D snoRNA.

The majority of research on box C/D snoRNA processing has been conducted using the yeast *S. cerevisiae*. Research in yeast has shown that the 5' extended sequences of several species of polycistronic and intronic box C/D pre-snoRNAs are processed by two homologous 5'-3' exonucleases, known as Rat1p and Xrn1p, which also function in

5' processing of the 5.8 rRNA (Petfalski et al, 1998; Figure 1.11). Further studies in yeast revealed that the 3' extended sequences of intronic box C/D pre-snoRNAs are processed by a 3'-5' exonuclease complex known as the exosome (Allmang et al, 1999; Figure 1.11). The exosome also functions in the processing of the 3' sequence of polycistronic box C/D snoRNAs once the extended sequences are first cleaved by the endonuclease Rnt1p (Kufel et al, 2000, 2003; Chanfreau, et al, 1998; Figure 1.11). Interestingly, Rnt1p has been also shown to function in the 3' cleavage of a number yeast intronic box C/D snoRNAs (Giorgi et al, 2001).

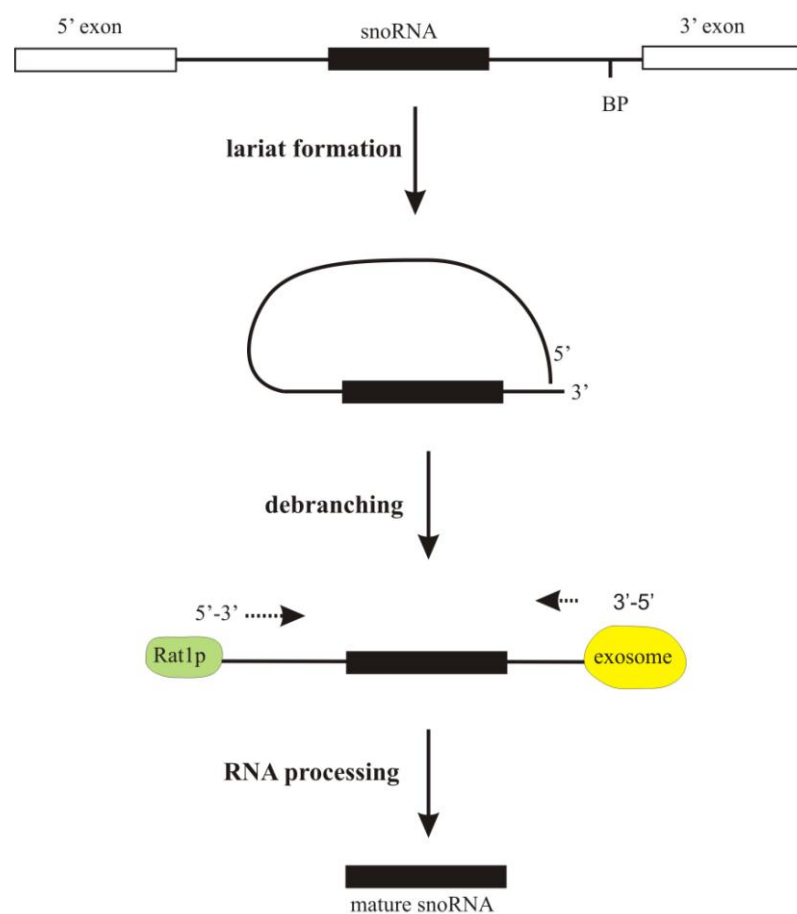


Figure 1.11: Processing of an intronic box C/D snoRNA in yeast

Exons are represented as white boxes while the box C/D snoRNA mature sequence is shown as a black box. The black line represents intronic mRNA sequences and pre-snoRNA sequences. The 5'-3' exonuclease Rat1p is shown as a green oval while the 3'-5' exonuclease exosome complex is represented as a yellow oval. Exonucleolytic processing is represented as black dashed arrows. Black arrows show progression of processing.

The exosome is a multiprotein complex found in both the nucleus and cytoplasm. Interestingly, the composition of the yeast exosome differs slightly between the nucleus and the cytoplasm; however, there are ten common components (Rrp41, Rrp46, Mtr3, Rrp42, Rrp43, Rrp45, Rrp4, Rrp40, Csl4 and Rrp44) (Mitchell et al, 1997; Allmang et al, 1999). In addition to the 3' processing of box C/D pre-snoRNAs the exosome has known functions in the processing snRNAs, mRNA degradation and the 3' processing of the 5.8S rRNA (Allmang et al, 1999, 2000; Anderson and Parker, 1998; Burkard et al, 2000; Torchet et al, 2002; Libri et al, 2002). The use of yeast mutants revealed that 3' processing of box C/D snoRNA requires the specific function of the Rrp6p component of the exosome (Allmang et al, 1999; Figure 1.11).

In yeast the processing of the box C/D pre-snoRNA 3' extended sequences is regulated by the Lhp1p protein (homologue of the human La protein), which has previously been shown to be essential for the stability of snRNAs during biogenesis (Xue et al, 2000; Pannone et al, 1998; Kufel et al, 2000; Figure 1.12). The Lsm (Like Sm) proteins are proposed to chaperone the association of the Lhp1p protein with the 3' snoRNA poly(U) sequences which then prevents excessive degradation of the 3' end of the snoRNA (Kufel et al, 2003).

The association of the box C/D snoRNP common core proteins with the snoRNA has also been implicated in regulating the processing of box C/D snoRNAs (Figure 1.12). It has been proposed that the association of core proteins displaces Lhp1p and defines the 3' end of the mature snoRNA (Kufel et al, 2003; Figure 1.12). Furthermore, analysis of the yeast U18 box C/D snoRNA revealed that the common core proteins are present throughout processing, and in particular the association of Nop1p (fibrillarin in humans) is a prerequisite for 3' cleavage of pre-snoRNA sequences by Rnt1p (Giorgi et al, 2001; Figure 1.12). Interestingly, Nop1p interacts directly with the Rnt1p, indicating that fibrillarin functions in the recruitment of this endonuclease (Giorgi et al, 2001).

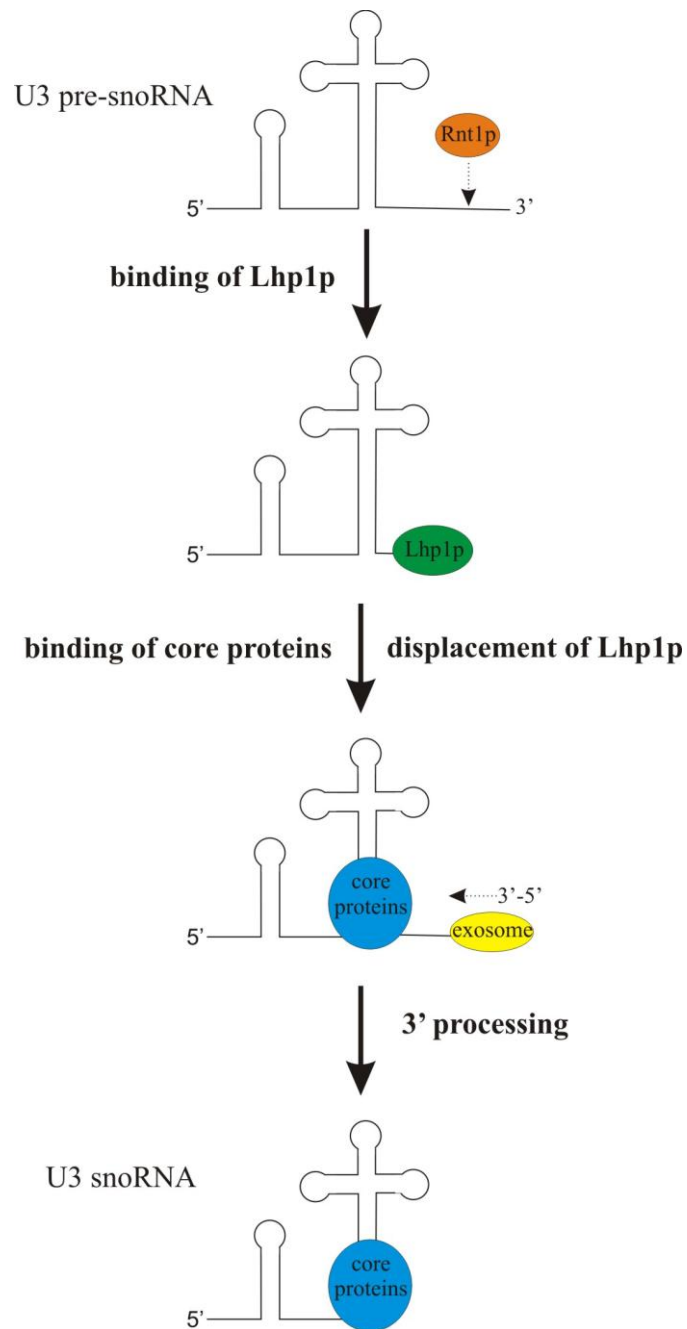


Figure 1.12: A model of 3' processing of the yeast U3 snoRNA

The U3 box C/D snoRNA is shown as a black line. The Rnt1p endonuclease is represented by a red oval while the Lhp1p protein is shown as a green oval. The box C/D snoRNP common core proteins are shown as a large blue circle. The 3' to 5' exonuclease exosome complex is represented as a yellow oval. Exo and endo nucleolytic activities are represented as black dashed arrows. Black arrows show progression of processing.

The processing of box C/D pre-snoRNAs in higher eukaryotes has not been as well characterised as in yeast; however, homologues of a number of yeast snoRNA processing factors also interact with the box C/D pre-snoRNP in human cells. For instance the La, Lsm4, Rrp46 (exosome) proteins associate with U3 and U8 box C/D pre-snoRNPs (Watkins et al, 2004; 2007). Analysis of the interactions of these factors with the U8 pre-snoRNA, of which there are four precursors (I to IV, with I being the longest), provided data on the order of binding of these factors (Watkins et al, 2007). The La protein specifically associates with pre-snoRNP complexes containing the U8 pre-snoRNA intermediate I while RNA processing factors such as Rrp46 and Lsm4 only associate later on in biogenesis with U8 pre-snoRNP complexes containing snoRNA intermediates III and IV (Watkins et al, 2007). Interestingly, the core proteins associate with the U8 pre-snoRNA intermediate I, which indicates that the core proteins bind early on in the biogenesis of the U8 snoRNP. Less is known about the 5' processing of box C/D pre-snoRNAs in higher eukaryotes with the only identified endonuclease being XendoU. This endonuclease was identified in *X. laevis* and is required for 5' cleavage events that release intronic U86 box C/D snoRNA (Laneve et al, 2002).

1.4.3. The 5' cap hypermethylation of box C/D snoRNAs

The 5' end of RNAs transcribed by RNA polymerase II, which includes mRNAs, snRNAs and independently transcribed snoRNAs, contain a post transcriptionally added 5' terminal 7-monomethylguanosine cap (m^7G). The 5' m^7G cap of snRNAs and snoRNAs is converted to a 2,2,7-trimethylguanosine cap (m_3G), which involves the addition of two methyl groups (Saponara and Enger, 1969).

The formation and the function of the 5' cap has been best characterised for snRNAs. In snRNA biogenesis the m^7G cap functions as a nuclear export signal resulting in the export of pre-snRNAs to the cytoplasm where further maturation occurs (Ohno et al, 2000). In the cytoplasm the m^7G cap of the snRNAs is converted, by the methyltransferase trimethylguanosine 1 (TGS1), to an m_3G cap that functions as nuclear import signal (Mouaikel et al, 2002).

The role of the m⁷G / m₃G cap of independently transcribed box C/D snoRNAs is not so clear, especially considering that the majority of human box C/D snoRNAs do not contain 5' caps as they are processed from either pre-mRNA or polycistrons with 5'-phosphate terminals (Tollervey and Kiss, 1997). The m₃G cap of independently transcribed snoRNAs could function either in cellular transport or protect the 5' end from RNA processing. While it is known that the snRNAs obtain their m₃G cap in the cytoplasm the cellular location where independently transcribed box snoRNAs acquire their m₃G cap is not clear with both the Cajal bodies and the cytoplasm implicated (Verheggen et al, 2002; Watkins et al, 2007).

Several studies have indicated that box C/D snoRNA maturation occurs exclusively in the nucleus and this includes the conversion of the 5' m⁷G cap to an m₃G cap (Terns and Dahlberg, 1994; Terns et al, 1995). Interestingly, two isoforms of TGS1 have been discovered in human cells and these alternatively localise to the cytoplasm and Cajal bodies (Girard et al, 2007). The longer 110kDa TGS1 preferentially localises to the cytoplasm while the smaller TGS1, which lacks the N terminal domain, is found exclusively in Cajal bodies (Girard et al 2007). The shorter isoform of TGS1 interacts with components of the box C/D snoRNPs, which indicates that it functions in modification of snoRNA (Girard et al, 2007). Furthermore, analysis in yeast revealed that both m⁷G and m₃G capped U3 snoRNA are present in the yeast equivalent of the Cajal body, known as the nucleolar body (Verheggen et al, 2002). This suggests that the m⁷G cap of independently transcribed box C/D snoRNAs is converted to an m₃G cap in the Cajal bodies.

A number of studies have indicated that box C/D snoRNAs are transported to the cytoplasm during biogenesis (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007). One study using *X. laevis* oocytes showed that the U3 box C/D snoRNA 5' m⁷G cap could be converted to an m₃G cap in the cytoplasm and that box C/D snoRNAs injected into the cytoplasm are imported into the nucleus (Baserga et al, 1992). Furthermore U3, U8 and U8 box C/D snoRNAs have been found in the cytoplasm of human cells (Watkins et al, 2007) and analysis of the maturation of the U8 snoRNA indicates that the m⁷G cap is converted to an m₃G cap in the cytoplasm (Watkins et al, 2007).

1.4.4. Role of assembly factors in box C/D snoRNP biogenesis

The box C/D snoRNP common core proteins associate with the snoRNA in a hierarchical manner. The binding of 15.5K to the box C/D snoRNA is essential for subsequent association of the other common core proteins (Watkins et al, 2002). It has been proposed that the binding of 15.5K to the K-turn of the box C/D motif causes a conformational change in the RNA, similar to the association of 15.5K with the U4 snRNA (Vidovic et al, 2000). This change in snoRNA confirmation has been proposed to create binding sites for NOP56, NOP58 and fibrillarin (Watkins et al, 2002).

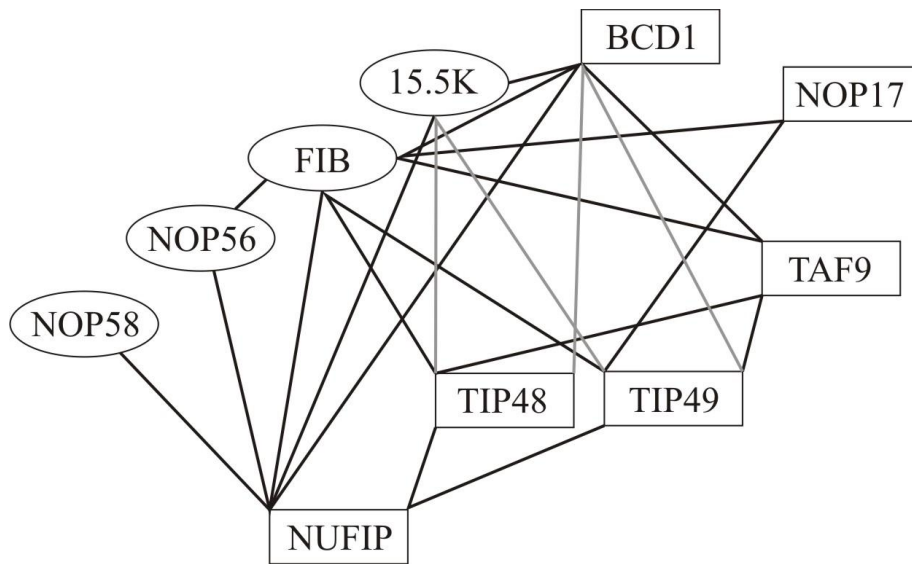


Figure 1.13: Schematic representation of the box C/D pre-snoRNP interaction network

The box C/D snoRNP core proteins are shown as ovals while the assembly factors are represented as rectangles. The black lines correspond to interactions between proteins. The grey lines represent interaction between proteins that are altered by either ATP or ADP. FIB corresponds to fibrillarin. Figure adapted from McKeegan et al, 2007.

A number of box C/D snoRNP assembly factors (TIP48, TIP49, BCD1, NOP17, NUFIP and TAF9) have been proposed to function in the early recruitment of the core proteins to the pre-snoRNP (McKeegan et al, 2007). These assembly factors associate with box C/D pre-snoRNP complexes containing pre-snoRNA and form an intricate interaction network with the common core proteins (Figure 1.13; Watkins et al, 2004; McKeegan et al, 2007).

It is proposed that during box C/D snoRNP biogenesis restructuring / remodelling events occur, possibly facilitated by the assembly factors, resulting in the stable association of the common core proteins with the snoRNA. This is supported by the fact that there are very few direct interactions between the core proteins themselves and that there is an increase in the stability of the association of both NOP56 and NOP58 between nucleoplasmic and nucleolar box C/D snoRNPs (Watkins et al, 2004; McKeegan et al, 2007).

1.4.4.1. TIP48 and TIP49

TIP48 and TIP49 (TBP (TATA binding protein) interacting protein of 48kDa or 49kDa) are a conserved class of AAA ATPases that interact with one another via Walker A and B motifs forming hexameric ring structures (Puri et al, 2007). The TIP48 and TIP49 proteins have functions in modifying chromatin structure (Shen et al, 2000) and in transcription activating complexes (Cho et al, 2001).

TIP48 and TIP49 have been shown to associate with nucleoplasmic box C/D pre-snoRNP complexes in both mouse and human cells (Newman et al, 2000; Watkins et al, 2004; 2007). The depletion of either TIP48 or TIP49, in human cells, results in the reduction of box C/D snoRNA levels which indicates that both proteins are required for box C/D snoRNP biogenesis (Watkins et al, 2004). *In vitro* interaction assays, using human proteins, showed that both TIP48 and TIP49 interact with the common core proteins 15.5K and fibrillarin, with the interaction with 15.5K dependent upon the presence of ATP (McKeegan et al 2007; Figure 1.13). TIP48 and TIP49 also interact with other box C/D snoRNP assembly factors, namely NUFIP, BCD1 and TAF9 and in addition TIP49 interacts with NOP17 (McKeegan et al, 2007; Figure 1.13).

Studies of the yeast homologues of TIP48 and TIP49 (known as Rvb1 and Rvb2, respectively) also indicate that these proteins function in box C/D snoRNP biogenesis. Mutation of the Walker A and B motifs of Rvb1 and Rvb2 result in the reduction of box C/D snoRNA levels suggesting that these motifs are vital for the function of Rvb1 and Rvb2 in box C/D snoRNP biogenesis (King et al, 2001). Furthermore, loss of Rvb2

results in the accumulation of fibrillarin outside the nucleolus indicating that Rvb2 functions in box C/D snoRNP biogenesis (King et al, 2001).

As TIP48 and TIP49 are the only assembly factors in the box C/D pre-snoRNP that are ATPases it has been proposed that these proteins function in the restructuring events that result in the stable association of the common core proteins to the snoRNA (Watkins et al, 2004). Interestingly, TIP49 associates with both U3 pre-snoRNA and mature snoRNA, although it is absent from nucleolar snoRNPs, which could indicate that TIP49 functions in the latter stages of box C/D snoRNP biogenesis perhaps in structural rearrangements (Watkins et al, 2004).

1.4.4.2. BCD1

The BCD1 protein (Box C/D snoRNA accumulation) was initially identified as a box C/D snoRNP interacting factor in a panoramic yeast two-hybrid screen of non-coding RNA processing proteins (Peng et al, 2003). Sequence analysis of BCD1 revealed the presence of a zinc finger domain, but no other notable structures or motifs.

The interaction of BCD1 with box C/D snoRNPs was further confirmed in yeast using TAP tag purification of BCD1, which resulted in the co-purification of box C/D snoRNA (Peng et al, 2003). The depletion of BCD1 in yeast results in the reduction of box C/D snoRNA levels suggesting that BCD1 is required for box C/D snoRNP biogenesis (Peng et al, 2003; Hiley et al, 2005).

Analysis of the human BCD1 homologue revealed that it associates with nucleoplasmic pre-snoRNPs and interacts with fibrillarin, 15.5K, NUFIP and TAF9 (Figure 1.13; McKeegan et al, 2007). The depletion of BCD1, in human cells, results in the reduction of box C/D snoRNA levels, which again suggests that BCD1 functions in box C/D snoRNP biogenesis (McKeegan et al, 2007).

1.4.4.3. NOP17

Nop17p (nucleolar localisation protein 17 also known as PIH1) was identified in a yeast two-hybrid screen as a protein that interacts with Nop58p and the exosome subunit Rrp43p (Gonzales et al, 2005). Further analysis in yeast revealed that Nop17p is required for box C/D snoRNA accumulation (Zhao et al, 2008) and for the correct cellular localisation of the box C/D snoRNP common core proteins (Gonzales et al, 2005).

Investigations into the human homologue of Nop17p (NOP17) showed that it is essential for maintaining box C/D snoRNA levels (McKeegan et al, 2007; Boulon et al, 2008). Further analysis revealed that NOP17 associates with nucleoplasmic pre-snoRNPs and interacts with fibrillarin, TIP48 and TIP49 (McKeegan et al, 2007; Figure 1.13).

1.4.4.4. NUFIP

NUFIP (Nuclear FMRP interacting protein) is a zinc finger containing transcription factor that interacts with the fragile X mental retardation protein (FMRP; Bardoni et al, 1999; Cabart et al, 2004). NUFIP was originally identified as a box C/D snoRNP interacting factor in a yeast two-hybrid screen (McKeegan et al, 2007; Boulon et al, 2008).

Analysis of human NUFIP revealed that it associates with nucleoplasmic box C/D pre-snoRNPs and interacts with all the box C/D snoRNP core proteins and assembly factors (Figure 1.13; McKeegan et al, 2007). NUFIP is also essential for the accumulation of box C/D snoRNA, which indicates that it functions in box C/D snoRNP biogenesis (Boulon et al, 2008).

The high number of interactions between NUFIP, the common core proteins and assembly factors indicates that NUFIP may function as a central scaffold or bridging protein in the box C/D pre-snoRNP. This potential function of NUFIP was investigated using an *in vitro* assembly assay and revealed that NUFIP could mediate the assembly

of a partial box C/D snoRNP consisting of NOP56, 15.5K and snoRNA (McKeegan et al, 2007). This provided functional evidence that NUFIP may bridge the interactions of the core proteins with the snoRNA in box C/D snoRNP biogenesis.

1.4.4.5. TAF9

TAF9 (transcription activating factor 9) is a TATA binding protein that functions in RNA polymerase II transcription (Milgrom et al, 2005). TAF9 was implicated in box C/D snoRNP biogenesis in a yeast two-hybrid screen, which revealed it interacts with NOP56 (Ito et al, 2001).

Investigations of human TAF9 showed that it associates with nucleoplasmic pre-snoRNPs (McKeegan et al, 2007) and interacts with the common core protein fibrillarin, but interestingly not NOP56, as well as the assembly factors BCD1, NUFIP, TIP48 and TIP49 (McKeegan et al, 2007; Figure 1.13).

1.4.5. Transport factors

The cellular localisation(s) where box C/D snoRNP biogenesis occurs is not clear, with evidence suggesting the nucleoplasm (Terns et al, 1995; Watkins et al, 2004; Boulon et al, 2004), Cajal bodies (Verheggen et al, 2001, 2002; Boulon et al, 2004; Lemm et al, 2006) and cytoplasm (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007). A variety of proteins known to function in the transport of other factors associate with the box C/D pre-snoRNP complexes and these, more than likely, target the pre-snoRNP to sub-cellular locations where biogenesis occurs (Yang et al, 2000; Boulon et al, 2004; Watkins et al, 2004, 2007).

1.4.5.1. Nuclear export factors

A number of snRNP nuclear export factors associate with box C/D pre-snoRNPs and these include CBC (Cap Binding complex), PHAX (phosphorylated adaptor for RNA export) and CRM1 (Chromosome maintenance region 1) (Boulon et al, 2004; Watkins et al, 2004, 2007). The role of these proteins in box C/D snoRNP biogenesis is not as

well defined as in snRNP biogenesis; however, they have been proposed to function in either intra-nuclear transport or nuclear export of box C/D pre-snoRNPs.

CBC, PHAX and CRM1 have all been shown to interact with U3 and U8 pre-snoRNA in the nucleoplasm of human cells (Boulon et al, 2004; Watkins et al, 2004, 2007). Analysis of PHAX revealed that it interacts with the m⁷G cap of U3 pre-snoRNAs in the absence of both CBC and CRM1 and has been proposed to function in the localisation of U3 snoRNA to the Cajal bodies as the over expression of U3 snoRNA results the accumulation of PHAX in the Cajal bodies (Boulon et al, 2004). However, PHAX has also been found associated with cytoplasmic U8 pre-snoRNP complexes which could indicate that it functions in the nuclear export of box C/D pre-snoRNPs (Watkins et al, 2007).

CRM1 has been found associated with nucleoplasmic U3 pre-snoRNAs that have an m₃G cap (Boulon et al, 2004). It has been proposed that CRM1 functions in the export of box C/D pre-snoRNAs from the Cajal bodies as inhibition of CRM1 prevents the localisation of labelled U3 snoRNA to the nucleolus and results in its accumulation in the Cajal bodies (Boulon et al, 2004). Furthermore both CRM1 (Boulon et al, 2004) and m₃G capped U3 pre-snoRNAs (Verheggen et al, 2002) are enriched in Cajal bodies.

1.4.5.2. Snurportin1

Snurportin1 has known functions in the nuclear import of m₃G capped snRNA (Yong et al, 2002) and has also been implicated in the biogenesis of box C/D snoRNPs as it has been found associated with cytoplasmic U8 pre-snoRNP complexes that contain m₃G capped snoRNA (Watkins et al, 2007). The depletion of Snurportin1, in human cells, results in the accumulation of U8 snoRNA in the cytoplasm and Cajal bodies (Watkins et al, 2007), which suggests that Snurportin1 functions in the nuclear import of cytoplasmic U8 pre-snoRNPs.

1.4.5.3. Nopp140

Nopp140 is a highly phosphorylated protein that shuttles between the nucleolus and cytoplasm in mammalian cells (Meier and Blobel, 1992). Nopp140 was identified as a potential box C/D snoRNP factor when it was shown to interact with fibrillarin and was found associated with nucleoplasmic box C/D pre-snoRNPs (Yang et al, 2000; Watkins et al, 2004). Furthermore, expression of a dominant negative form of Nopp140 results in the accumulation of fibrillarin outside of the nucleolus (Yang et al, 2000). However, analysis in yeast showed that the Nopp140 homologue (Srp40p) was only essential for maintaining box H/ACA and not box C/D snoRNA levels (Yang et al, 2000). This could indicate the Nopp140 has a more vital for the biogenesis of the box H/ACA snoRNPs than the box C/D snoRNPs.

1.4.6. Sites of box C/D snoRNP biogenesis

Box C/D snoRNP biogenesis occurs, at least in part, in the nucleoplasm where the snoRNA is transcribed as an extended precursor by RNA polymerase II (Figure 1.14). The processing of the pre-snoRNA extended sequences by endo and exo nucleases has been shown to be mediated / regulated by the association of the common core proteins (Giorgi et al, 2001) and proteins such as La (Kufel et al, 2003; Watkins et al, 2004; 2007). The association of the common core proteins with the snoRNA has been proposed to be facilitated by the assembly factors that associate with the nucleoplasmic pre-snoRNP complexes (Watkins et al, 2004, 2007; McKeegan et al, 2007; Figure 1.14).

The Cajal bodies have also been implicated in box C/D snoRNP biogenesis as not only have box C/D snoRNA genes have been found associated with the Cajal bodies (Gao et al, 1997) but these nuclear bodies are enriched in both snoRNA and the common core proteins (Verheggen et al, 2001). The Cajal bodies have been proposed as the site of the association of the core proteins (Verheggen et al, 2001), and also the site where independently transcribed snoRNAs obtain their m₃G cap (Verheggen et al, 2002; Boulon et al, 2004; Figure 1.14). However, as of yet, there is no direct evidence that either core protein association or m₃G cap formation occurs in this nuclear body.

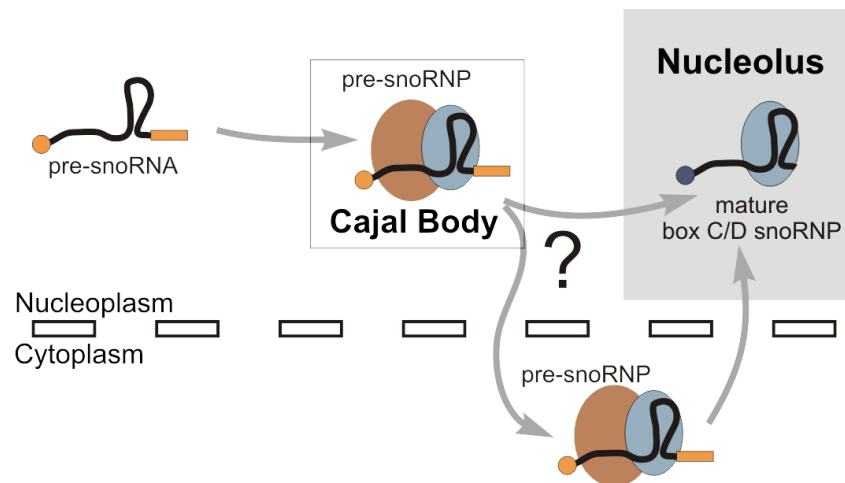


Figure 1.14: Box C/D snoRNP biogenesis

A black line represents the box C/D snoRNA with the 5' m⁷G cap shown as an orange circle and the 3' extended sequence as an orange rectangle. The core proteins are represented by a blue oval. The assembly, RNA processing and transport factors are represented by a brown oval. The 5' m₃G cap of mature snoRNA is shown as by a dark blue circle. The Cajal body is represented as a grey outlined box while the nucleolus is shown as a solid grey box. Grey arrows show the progression of biogenesis. The nuclear membrane is represented by black outlined rectangles. Image provided by N.J.Watkins, Newcastle University, UK.

Several studies have indicated that box C/D snoRNA biogenesis occurs exclusively in the nucleus (Terns and Dahlberg, 1994; Terns et al, 1995); however, there is evidence that box C/D snoRNP biogenesis occurs, in part, in the cytoplasm (Figure 1.14). For instance not only do the U3, U8 and U13 pre-snoRNPs interact with the snRNP nuclear export proteins PHAX, CBC, CRM1 and Ran (Boulon et al, 2004; Watkins et al, 2004, 2007) but U3, U8 and U13 snoRNAs have been found in cytoplasmic extracts (Watkins et al, 2007). Furthermore, investigation of cytoplasmic U8 pre-snoRNPs revealed two complexes: one pre-snoRNP containing 5' m⁷G capped snoRNA and the nuclear export factor PHAX and another pre-snoRNP complex containing 5' m₃G capped snoRNA and the nuclear import factor Snurportin1 (Watkins et al, 2007). This could indicate the PHAX exports the U8 pre-snoRNP to the cytoplasm where the m⁷G cap is converted to an m₃G cap and then the pre-snoRNP associates with Snurportin1 and is transported back to the nucleus.

1.5. Small nuclear ribonucleic protein complexes (snRNPs)

The small nuclear RNPs (snRNPs) function in pre-mRNA splicing as part of a larger RNP complex known as the spliceosome. The snRNPs consist of core proteins and a ncRNA, known as the snRNA, which recognises intronic sequences and is essential for processes that remove the intron.

1.5.1. SnRNP structure

The U1, U2, U4, U5 and U6 snRNPs are responsible for the majority of pre-mRNA splicing. The vertebrate snRNAs are between 100 and 200 nucleotides long and contain a conserved site Sm site (with the exception of U6) that has the consensus sequence RAUUUUUGR (where R is purine). The Sm proteins (B/B', D3, D2, D1, E, F and G) associate with the Sm site in a heptameric ring structure known as the Sm core. The snRNAs also contain stem loop structures that bind snRNP type specific core proteins (Figure 1.15).

The best structurally characterised snRNP is U1 which contains a 164 nucleotide long snRNA that forms a twin hairpin structure consisting of four stem loops (I, II, III and IV; Figure 1.15). The Sm core of the U1 snRNP is located between the hairpin structure while the U1-70K and U1-A core proteins interact with stem loops I and II, respectively. The U1-C core protein associates with the U1 snRNP through an interaction with the U1-70K protein (Oubridge et al, 1994; Stark et al, 2001; Figure 1.15).

The U6 snRNP is distinctive from the other snRNPs as the snRNA does not contain an Sm site and does not associate with Sm proteins. The U6 snRNAs associate with the Lsm proteins (2, 3, 4, 5, 6, 7 and 8) that form a heptameric ring structure at a 3' uridine rich sequence (Achsel et al, 1999; Vidal et al, 1999). The Lsm proteins contain the Sm motif (Hermann et al, 1995) and have been proposed to function in the formation of the U4/U6.U5 tri-snRNP (Achsek et al, 1999) and also the regulation of the 3' processing of some RNAs (Kufel et al, 2003). Unsurprisingly, the U6 snRNPs undergo a different biogenesis pathway to the other snRNPs.

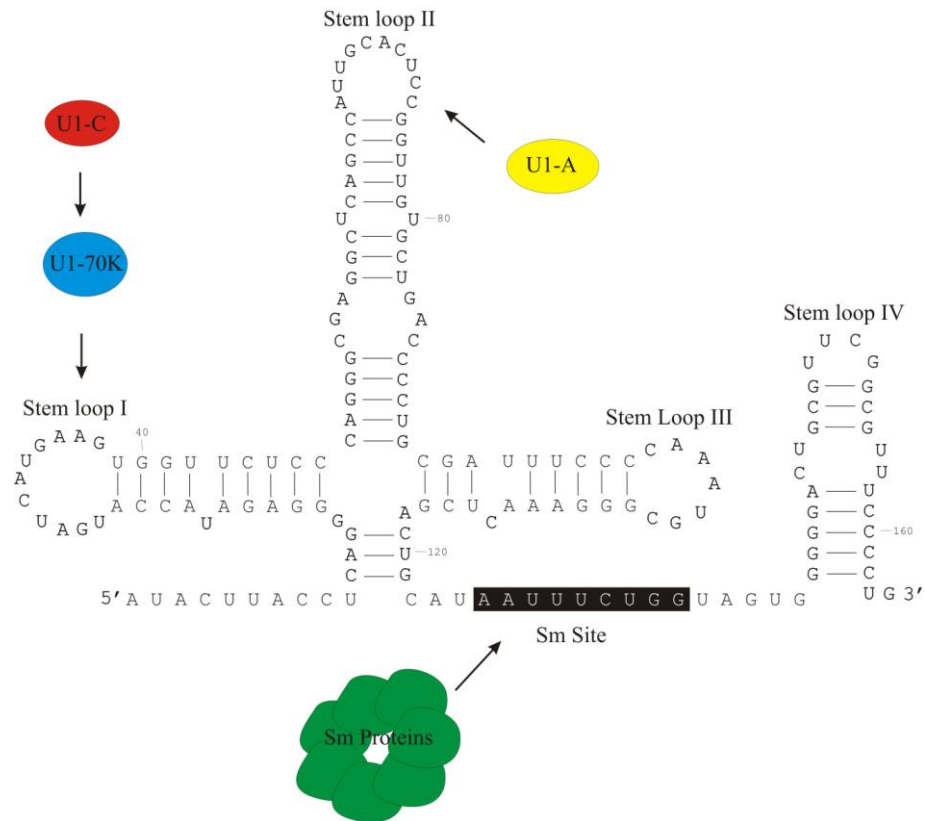


Figure 1.15: Human U1 snRNP structure

The snRNA sequence is represented as black text except for the conserved Sm site, which is shown as white text on black. Numerical values alongside snRNA sequence correspond to nucleotide number. The Sm proteins are represented by a green ovals and the U1-A protein is shown as a yellow oval. The U1-C and U1-70K proteins are represented as red and blue ovals, respectively.

The U4, U5 and U6 snRNPs function in pre-mRNA splicing as part of a U4/U6.U5 tri-snRNP complex. The formation of the U4/U6.U5 tri-snRNP complex is a hierarchical process that occurs in the Cajal bodies (Stanek et al, 2003; Schaffert et al, 2004). The formation of the U4/U6.U5 tri-snRNP first requires the association of 15.5K to the 5' stem loop of the U4 snRNA, which stabilizes a K-turn structure, followed by the binding of hPRP31 (61K) and a number of other core proteins (Nottrott et al, 1999; Vidovic et al, 2000). A U4/U6 di-snRNP is then formed, which involves the p110 protein (Stanek et al, 2003), followed by the association of the CypH (20K), hPRP4 (60K) and hPRP3 (90K) core proteins (Nottrott et al, 1999; 2002). The U5 snRNP is incorporated into the U4/U6.U5 tri-snRNP through an interaction between the U5 specific hPRP6 (102K) and the U4 specific hPRP31 protein (Makarova et al, 2002).

1.5.2. Pre-mRNA splicing and snRNPs

Pre-mRNA splicing is performed by the spliceosome which is formed by the ordered, stepwise association of snRNPs and splicing factors to the pre-mRNA in what are termed the E, A, B and C complexes.

Assembly of the spliceosome begins with the formation of the E complex, which involves the binding of the U1 snRNP to the 5' splice site and the association of the branch point binding protein (BBP) and U2 auxiliary factor (U2AF) to the branch site and 3' pyrimidine tract, respectively (Figure 1.16). The A complex is formed by the association of the U2 snRNP to the branch point site through interactions with U2AF and BBP, which are subsequently displaced (Will and Luhrmann, 1997).

Prior to associating with the spliceosome the U4, U6 and U5 snRNPs are preassembled into the U4/U6.U5 tri-snRNP in Cajal bodies (Schaffert et al, 2004). In the tri-snRNP the U4 and U6 snRNAs interact stably through direct base pairing while U5 loosely associates through interactions between core proteins. The U4/U6.U5 tri-snRNP interacts with the spliceosome at the 5' splice site (Maroney et al, 2000) forming the B complex, and results in numerous structural changes creating a complex RNA-RNA interaction network between the pre-mRNA and the snRNPs (Figure 1.16).

The structural changes that occur in the formation of complex B cause the displacement of the U1 and U4 snRNPs resulting in the U6 and U5 snRNPs directly interacting with the U2 snRNP (Hastings and Krainer 2001), forming the C complex. These rearrangements bring the 5' splice site and the branch point into close proximity, which in conjunction with other splicing factors results in the release of the intronic lariat and ligation of the exons (Figure 1.16). The final enzymatic processes that result in the excision of the lariat is not fully understood; however, it is possible that the snRNAs are themselves the catalysts forming an active site / ribozyme in coordination with metal ions (Yean et al, 2000).

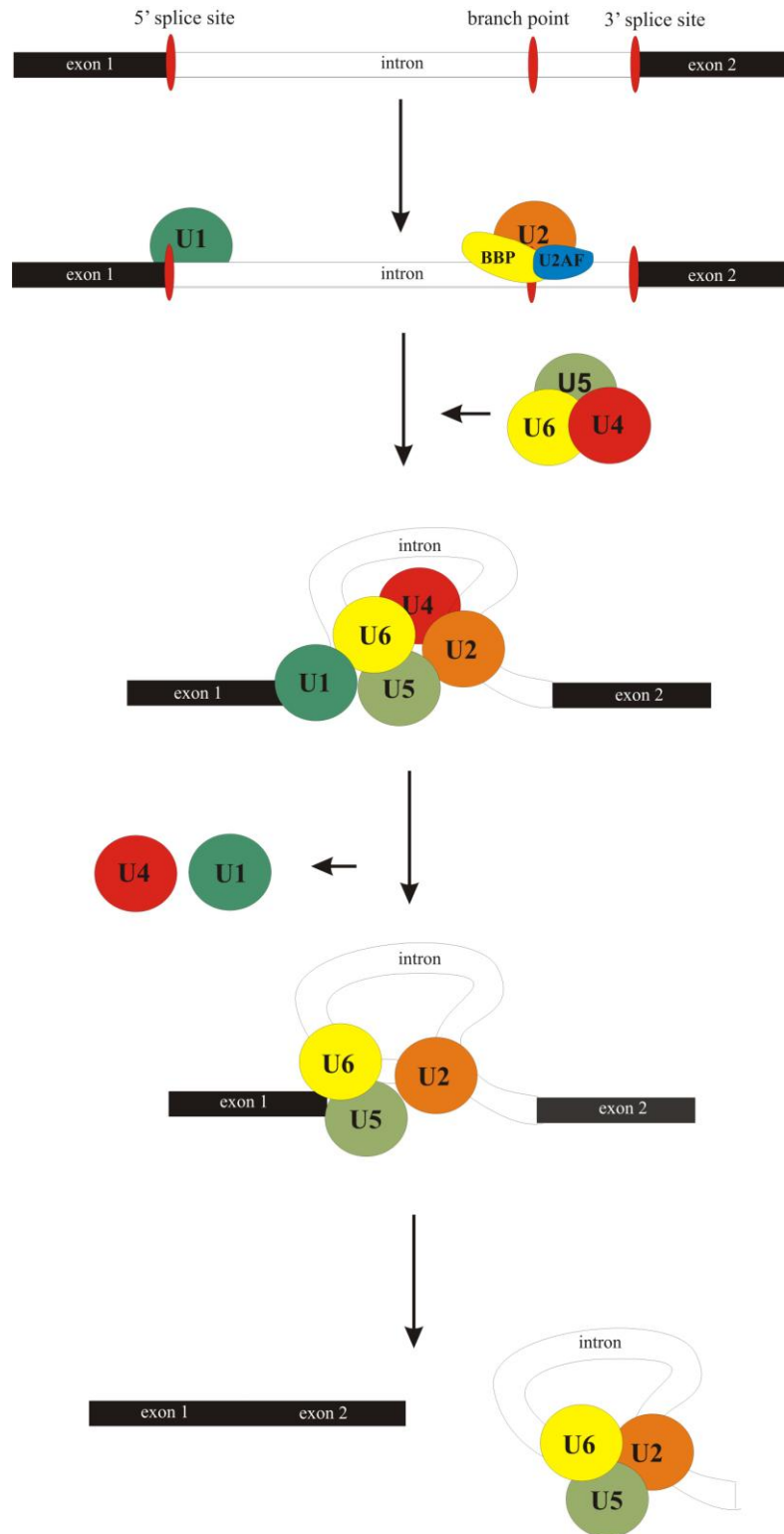


Figure 1.16: Pre-mRNA splicing

The exons are shown as black rectangles while the intron is represented as a white rectangle. The 3' splice site, branch point and 5' splice site are shown as red ovals. The branch point protein (BBP) and U2AF are represented by yellow and blue shapes, respectively. The U1 (green), U2 (orange), U4 (red), U6 (yellow) and U5 snRNPs (pale green) are represented as coloured circles. The arrows indicate progression of splicing.

1.5.3. snRNP Biogenesis

The biogenesis of the U1, U2, U4 and U5 snRNPs is a tightly orchestrated and regulated process that occurs in both the cytoplasm and nucleoplasm utilising numerous assembly, RNA processing and transport factors (Will and Luhrmann, 2001; Patel and Bellini, 2008; Figure 1.17).

1.5.3.1. Transcription and nuclear export

The U1, U2, U4 and U5 snRNAs are transcribed by RNA polymerase II in the nucleoplasm from independent genes (Hernandez, 2001). The snRNAs contain their own independent promoters with a PSE, which functions as a TATA box, and a DSE, which functions as an enhancer. The transcribed pre-snRNAs contain a 5' m⁷G cap and an extended 3' sequence (Medlin et al, 2003).

In the nucleoplasm pre-snRNAs are subjected to exonucleolytic processing, which results in the production of more stable pre-snRNAs containing a small five to ten nucleotide 3' extension (Medlin et al, 2003). The 5' m⁷G cap of pre-snRNAs functions as an export signal and is bound by CBP, which recruits PHAX, and in turn binds the export receptor CRM1 to form pre-snRNP export complexes (Ohno et al, 2000; Segref et al, 2001). The PHAX associated with the pre-snRNP export complexes is phosphorylated by CK2 kinase and recruits RanGTP resulting in the export of the complexes via the nuclear pore complex (NPC; Figure 1.17). In the cytoplasm PHAX is dephosphorylated by protein phosphatase 2A and GTP hydrolysis of Ran causes the disassembly of the export complexes (Ohno et al, 2000; Kitao et al, 2008). Once disassembly has occurred the protein components of the export complex are recycled back to the nucleus.

1.5.3.2. Cytoplasmic modifications

In the cytoplasm the Sm proteins (B/B', D1, D2, D3, E, F and G) associate with the Sm site of the snRNAs forming a heptameric Sm core. The formation of the Sm core is

facilitated by the survival of motor neurons (SMN) complex (Paushkin et al, 2002; Figure 1.17).

The SMN complex binds preformed Sm sub-complexes (D3/B, D1/D2 and E/F/G), which are associated with pICln (a component of PRMT5 (protein arginine methyltransferase 5) complex), and then catalyses the formation of the Sm core on the pre-snRNAs (Chari et al, 2008). The SMN complex consists of the SMN, Gemin2-8 and UNRIP (unr (upstream of N-ras) interacting protein) factors, several of which interact with the Sm proteins. The SMN protein directly binds the B, D1 and D3 Sm proteins (Friesen et al, 2001; Brahms et al, 2001), Gemin3 interacts with the B, D2 and D3 Sm proteins (Charroux et al, 1999) and Gemin4 associates with the B, D1, D3 and E Sm proteins (Charroux et al, 2000). Gemin5 interact with the B, D1, D2, D3 and E Sm proteins (Gubitz et al, 2002), Gemin6 binds to the D2, D3, B and E Sm proteins (Pellizzoni et al, 2002) and Gemin7 interacts with the B, D2, D3 and E Sm proteins (Baccon et al, 2002; Carissimi et al, 2005).

In vitro assembly assays, using human cell extracts, showed that SMN and Gemin2, 3, 4 and 6 are essential for Sm core formation (Sphargel and Matera, 2005; Feng et al, 2005). Recently, however, a minimal complex consisting of just SMN and Gemin2 has been shown to be sufficient to facilitate Sm core formation (Kroiss et al, 2008, Chari et al, 2008). This could imply that either some of the Gemin proteins have redundant functions in Sm core assembly or that they have roles outside of snRNP biogenesis.

As the Sm proteins are able to spontaneously assemble on snRNAs *in vitro* (Raker et al, 1999) the SMN complex, more than likely, functions as either an enhancer of the kinetics of Sm core formation or as a regulatory protein ensuring that Sm cores are only assembled on snRNAs that fulfil a certain criteria. The SMN complex could function in a regulatory role as it binds to a specific site (5'-AUUUUUG-3') on the snRNAs, a certain distance away from the Sm binding site (Yong et al, 2004). Using these markers of spatial orientation the SMN complex could determine which snRNAs are suitable for Sm core formation. The SMN complex may also thwart spontaneous association of Sm proteins on deleterious snRNAs by preventing the formation of a free pool of Sm core proteins. Interestingly, Gemin5 is able to interact with the Sm site and parts of adjacent stem loops structures of snRNAs independently of the SMN complex (Battle et al,

2006). This could indicate that Gemin5 catalyses the formation of the SMN complex on the snRNAs.

After the formation of the Sm core the snRNAs 5' m⁷G caps are converted to m₃G caps by TGS1 (Figure 1.17). Interestingly, SMN has been shown to interact with TGS1, which could indicate that SMN recruits TGS1 to the snRNAs (Mouaikel et al, 2003). Finally in the cytoplasm the 3' extended sequences of the pre-snRNAs are trimmed by yet unidentified factors in metazoans (Wieben et al, 1985). However, in yeast pre-snRNAs have been shown to be processed by the 3'-5' cytoplasmic exosome (Allmang et al, 1999).

1.5.3.3. Nuclear import

The nuclear import of snRNPs can occur through two pathways, which either use the 5' m₃G cap (m₃G cap dependent pathway) or the Sm core as a nuclear localisation signal (NLS; Sm dependent pathway). Despite the different NLS both pathways utilise the nuclear import receptor importin β (Figure 1.17).

In the m₃G cap dependent pathway the nuclear import factor Snurportin1 binds the 5' m₃G cap of the snRNAs, via its C terminal domain, and interacts with the import receptor importin β, with its N terminal domain (Huber et al, 1998). SMN has been proposed to function in the Sm dependent pathway as not only does the SMN complex facilitate the formation of the Sm cores but it also interacts with importin β (Narayanan et al, 2002). Furthermore, SMN and importin β have been shown to be sufficient for snRNP nuclear import (Narayanan et al, 2002; Ospina et al, 2005). The pre-snRNP import complexes containing both SMN and Snurportin1 (Narayanan et al, 2002) are transported to nucleus through a NPC (Rollenhagen et al, 2003). The dissociation of an importin β complex in the nucleoplasm usually requires RanGTP hydrolysis; however, this is not the case in snRNP import possibly because the Ran binding site on importin β is partially obstructed by Snurportin1 (Vetter et al, 1999; Wohlwend et al, 2007).

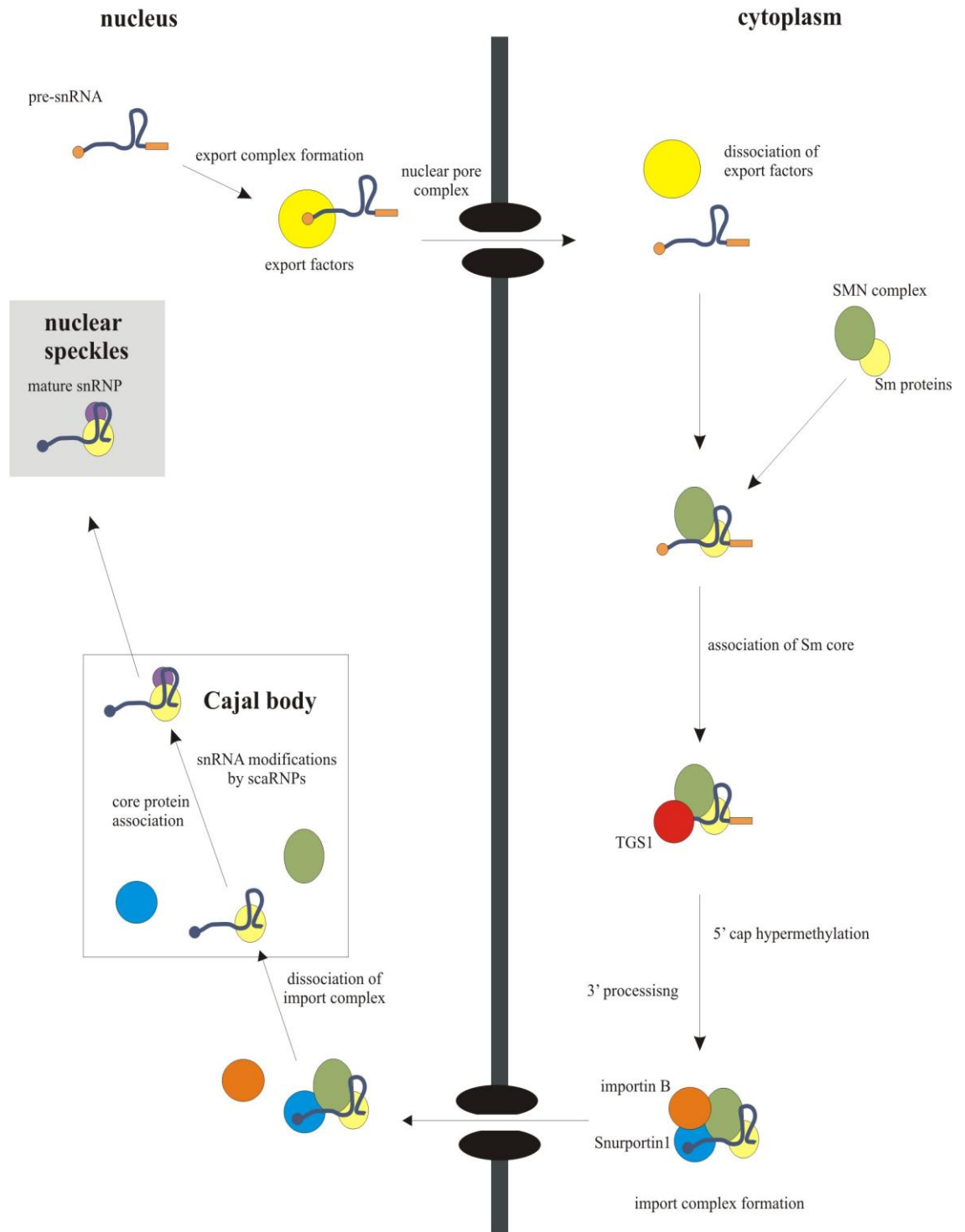


Figure 1.17: U1, U2, U4 and U5 snRNP biogenesis

The snRNA is shown as a grey line with the 5' m⁷G cap shown as an orange circle and the 3' extended sequence as an orange rectangle. Core proteins and assembly, transport and RNA processing factors are represented as coloured shapes. The 5' m₃G cap of mature snRNA is shown as by a dark grey circle. The Cajal body is represented as a grey outlined box while the nuclear speckle is shown as a solid grey box. Nuclear membrane is shown as a thick black line. Nuclear pore complex is represented as black ovals. Key steps of biogenesis are indicated. Arrows represent progression of biogenesis.

1.5.3.4. Nuclear maturation

Once the pre-snRNPs have entered the nucleus they localise to the Cajal bodies where the scaRNPs perform pseudouridylation and 2'-*O*-methylation of target snRNA residues (Jady et al, 2003; Figure 1.17). These chemical modifications are essential for the function of the snRNPs, for instance the chemical modifications within the first 20 nucleotides of the U2 snRNA are vital for the formation of the spliceosomal E complex (Donmez et al, 2004).

The Cajal bodies have been implicated as the site where the snRNP core proteins are associated and the U4/U6 di and U4/U6.U5 tri-snRNPs are formed (Stanek et al, 2003; Nesic et al, 2004; Schaffert et al, 2004). Once the final maturation processes in the Cajal bodies are complete the mature snRNPs localise to nuclear speckles where they are stored until they are required for pre-mRNA splicing (Lamond and Spector, 2003; Figure 1.17).

1.5.4. U6 snRNP biogenesis

The U6 snRNA is transcribed from an independent gene with a 3' terminal stretch of uridines by RNA polymerase III. The mature U6 snRNA contains a 5' γ -monomethylphosphate (mpppG) cap; however, it is not known at what stage or where this cap is formed (Singh and Reddy, 1989).

Unlike the U1, U2, U4 and U5 snRNPs the biogenesis of U6 snRNP is believed to be an entirely nuclear process. The 3' uridine sequence of U6 snRNA is initially bound by the La protein, which stabilises the snRNA (Wolin and Cadervall, 2002). A heptameric ring of Lsm proteins then replaces the La protein but the mechanisms that facilitate its formation are not known (Achsel et al, 1999). The U6 snRNA also undergoes 2'-*O*-methylation and pseudouridylation at specific residues; however, unlike the other snRNPs these events are performed by the nucleolar snoRNPs (Tycowski et al, 1998; Ganot et al, 1999). The mature U6 snRNPs localise to the Cajal bodies where they are formed into the U4/U6 di and U4/U6.U5 tri-snRNPs (Stanek et al, 2003; Nesic et al, 2004; Schaffert et al, 2004) prior to localising to the nuclear speckles.

1.6. SMN and spinal muscular atrophy (SMN)

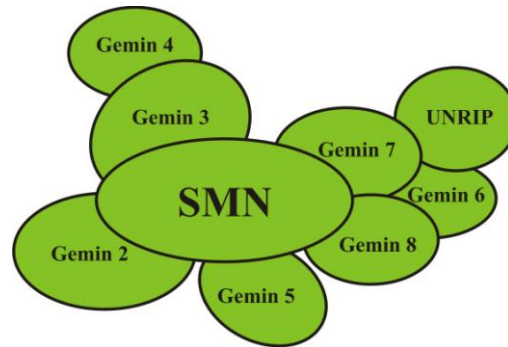
The SMN complex functions in the assembly of the snRNPs and has also been implicated in the biogenesis of telomerase RNPs (Bachand et al, 2002), miRNPs (Mourelatos et al, 2002) and snoRNPs (Jones et al, 2001; Pellizzoni et al, 2001; Lemm et al, 2006). The SMN complex has been under intense investigation as mutation of the gene encoding the SMN1 protein is linked to the neurodegenerative disease SMA.

1.6.1. The SMN complex

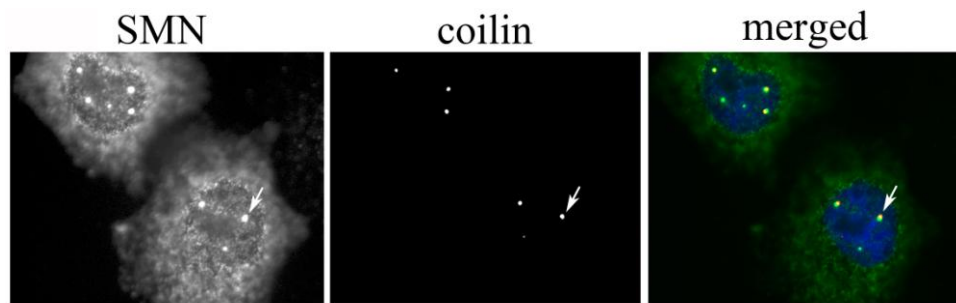
The SMN complex consists of SMN, Gemins2-8 and UNRIP. The Gemins were named based on the observation that they co-localise with SMN in nuclear Gems and Cajal bodies. The SMN complex is found diffusely in the cytoplasm and enriched in Cajal bodies and nuclear Gems (Figure 1.18 B; Liu and Dreyfuss, 1996). Interestingly, only cytoplasmic SMN complexes contain the UNRIP factor (Carissimi et al, 2005).

The SMN complex is held together through direct interactions between the protein components. SMN interacts with Gemin2, 3, 5, 7 and 8 while Gemin4 and 6 associate with the complex through interactions with Gemin3 and 7. Gemin8 also interacts with Gemin6 and 7. The UNRIP protein interacts with Gemin6 and 7 (Figure 1.18 A). The exact stoichiometry of the SMN complex is unknown and it has been proposed that the SMN protein self oligomerises resulting in the formation of larger complexes (Paushkin et al, 2002).

A



B



C

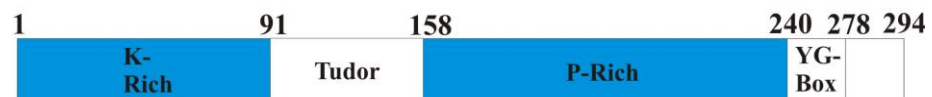


Figure 1.18: SMN structure and localisation

(A) Schematic diagram of the SMN complex. The SMN complex proteins are shown as green ovals.

(B) Localisation of SMN. HeLa SS6 cells were transfected with a plasmid expressing the SMN protein fused to an N terminal FLAG tag. Immunofluorescence was performed using anti-FLAG antibodies to detect tagged SMN. Anti-coilin antibodies were used to identify the Cajal bodies and DAPI (4,6-diamidino-2-phenylindol) staining of DNA used to show nucleoplasm. Merged image of shows SMN (green), coilin (red) and DAPI (blue). White arrow indicates a nuclear Gem associated with a Cajal body. The bar to the bottom right of the panel represents 5 μ m (micrometers).

(C) Schematic of SMN domain organisation. SMN contains a central tudor domain flanked by a N terminal lysine (K) rich sequence, C terminal proline (P) rich sequence, and a C terminal domain YG-box (tyrosine glycine). Numerical values above panel indicate amino acid number.

The SMN protein contains a number of conserved motifs such as an N terminal domain K-rich (lysine), which is a characteristic RNA binding motif (Figure 1.18 C; Bertrand et al, 1999), and a central tudor domain, which mediates interactions with the RGG motifs of numerous proteins including coilin and the Sm proteins (Selenko et al, 2001; Whitehead et al, 2002). The C terminal domain of SMN contains a P-rich (proline) sequence that binds the protein profilin, which functions in actin dynamics (Giesemann et al, 1999), and an YG-box (tyrosine and glycine) which is potentially involved in SMN self association and oligomerisation (Lorson et al, 1998).

A number of the other SMN complex proteins also contain conserved motifs, for instance Gemin3 contains a DEAD box RNA helicase motif, which are known to function in RNA metabolism (Charroux et al, 1999). Gemin5 houses 13 WD repeats in the N terminal domain, which are implicated in coordinating multi-protein complex assemblies, and coiled-coil motifs in the C-terminal domain, that have known roles in protein dimerization (Gubitz et al, 2002). Furthermore, Gemin5 has been shown to associate with snRNAs independently of the SMN complex (Battle et al, 2006). Gemin7 contains several RGG motifs that are required for the interaction with SMN (Baccon et al, 2002). The UNRIP factor contains WD repeats, which are implicated in coordinating multiprotein complex assemblies (Carissimi et al, 2005). Gemin2, 4, 6 and 8 have no obvious protein motifs (Liu et al, 1997; Charroux et al, 2000; Pellizzoni et al, 2002).

1.6.2. Genetics of SMA

In SMA the anterior horn alpha motor neurons progressively die resulting in atrophy of the surrounding muscle due to lack of nervous stimulation. SMA is an inherited disorder with around 1 in 6000 children affected and frequently results in early infant mortality (Schmalbruch and Haase, 2001).

The SMA locus maps to an inverted 500 kb repeat on chromosome 5q13, which contains the SMN1 gene (Wirth, 2000). The cause of SMA in 96 % of cases is a mutation that results in the loss of expression of SMN1 (Wirth, 2000). Interestingly, humans (and other higher primates) have a second centromeric copy of SMN1 known as SMN2 that only differs by five nucleotides. However, a C>T transition at position + 6

of exon 7 (840) of SMN2, which is within an exon splicing region, causes the frequent skipping of exon7 resulting in the population of expressed SMN2 being around 10 % full length and 90 % SMN Δ exon7 (Gavrilov et al, 1998; Soler-Botija et al, 2005). Despite the loss of exon7 SMN2 appears to be able to compensate for the loss of SMN1 and allows the survival of some SMN1-deleted patients. In SMA patients the number of SMN2 genes correlates with the severity of the disease, for example patients with type 1 SMA (the most severe form) have one or two copies of SMN2, patients with less severe SMA type II have three or four copies of SMN2 while SMA patients carrying five or six copies of SMN2 are only moderately affected with the disease (Prior et al, 2004; Wirth et al, 2006). The ability of SMN2 to compensate for the lack of SMN1 has also shown in a SMA mouse model (Monani et al, 2000). A minority of SMA patients (4 %) have mutations in SMN1 that do not prevent its expression but rather disturb the function of expressed SMN1 and, so far, around forty of these mutations have been characterised (Wirth et al, 1999; Bussaglia et al, 1995; Sun et al, 2005).

The underlying molecular mechanisms that cause the death of alpha neurons and result in SMA are not known. Several possibilities have been investigated but it remains unclear how the mutation of SMN1, which is required in all cells types, leads to the specific degeneration of anterior horn alpha motor neurons. A number of SMN functions have been analysed in the context of SMA and these include snRNP biogenesis (Zhang et al, 2008), outgrowth and pathfinding of neuritis (McWhorter et al, 2003), formation of the neuromuscular junction (Chan et al, 2003), axonal RNA transport (Rossoll et al, 2003) and muscle specific SMN roles (Rajendra et al, 2007).

1.6.3. SMN, SMA and snRNP biogenesis

The SMN complex is involved in the biogenesis of the spliceosomal snRNPs where it functions in Sm core assembly and in processes required for snRNP nuclear import (Yong et al, 2004). It is plausible that that dysfunction of snRNP biogenesis could lead to SMA due to issues of pre-mRNA splicing.

Experiments in animal models have shown that the injection of purified snRNPs into SMN depleted *X. laevis* and zebrafish embryos is able to prevent the developmental

arrest and motor neuron degeneration of the respective embryos (Winkler et al, 2005). Investigations into the effect of the SMN mutations on snRNP biogenesis in human cells have indicated that binding of Sm core proteins with the snRNAs is disrupted (Shpargel and Matera, 2005). For instance SMN lacking exon7 (SMN2) or containing SMA type I mutations (I116F, E134K, Q136E, and Y272C) are unable to support snRNP formation using *in vitro* assembly assays (Shpargel and Matera, 2005). It also appears that the localisation of snRNPs is affected in environments expressing the mutant SMN proteins. For example, in cells expressing SMN lacking exon7 snRNPs fail to localise to Cajal bodies (Shpargel and Matera, 2005), a phenotype that is seen in SMA patient fibroblast cells (Renvoise et al, 2006). The E134K (tudor domain) and Y272C (YG box) SMA causing mutations of SMN have also been shown to prevent the nuclear import of SMN and snRNPs (Narayanan et al, 2004).

Analysis of tissues from a SMA mouse model has shown that alterations in pre-mRNA splicing are not just occurring in alpha motor neurons but in all cell types (Zhang et al, 2008). What is particularly interesting is that in different cell types, from the SMA mouse model, different pre-mRNAs are affected as well as the repertoires of snRNPs (Zhang et al, 2008). A possible explanation for the cell type specific changes in pre-mRNA splicing is that as each cell type has different sets of splicing factors and the loss of SMN results in a unique change in the cellular repertoires of snRNPs, this results in cell type specific alterations of pre-mRNA splicing. This differential effect of the loss of SMN provides evidence of how the loss or dysfunction of a ubiquitous protein can cause tissue specific changes and in the case of SMN the specific death of alpha motor neurons in the anterior horn.

1.6.4. SMN, SMA and mRNA trafficking

SMN has been proposed to function in the intracellular trafficking of mRNA in motor neurons. Immunocytochemical analysis of neurites and growth cones of cultured neurone cells revealed that SMN is enriched in granules that possess bi directional cytoskeletal movements consistent with mRNA transport (Zhang et al, 2003; Bassell and Singer, 2001). Further evidence of SMN functioning in mRNA transport in neurons came from the observation that SMN interacts with two related heterogeneous nuclear

RNPs (hnRNP) R and Q that bind mRNA (Rossoll et al, 2002). Interestingly, hnRNP R was shown to interact with β -actin mRNA which requires localisation to the growth cones and is proposed to be involved in axonal growth. Investigations into the relationship between SMN and the hnRNPs revealed that the loss of SMN causes a disruption of hnRNP localisation and the subsequent loss of β -actin mRNA from growth cones (Rossoll et al, 2003). As the loss of SMN leads to the subsequent loss of β -actin mRNA at growth cones this could lead to reduced axonal growth and SMA.

1.6.5. SMN, SMA and muscle specific functions

Analysis of a *Drosophila melanogaster* SMA model indicated that dysfunction of a muscle specific function of SMN may cause SMA (Rajendra et al, 2007). This SMA model revealed that in depleted SMN environments the expression of the muscle specific actin isoform actin88F is reduced in the absence of any pre-mRNA splicing defect (Rajendra et al, 2007). SMN has also been shown to directly interact with two muscle proteins, α -actinin and $\alpha\beta$ -crystallin (Rajendra et al, 2007). Furthermore, analysis of skeletal muscle tissue from SMN knockout mice and human SMA type III patients with no neuronal degeneration revealed dystrophic muscle phenotypes (Cifuentes-Diaz et al, 2001; Muqit et al, 2004).

1.7. SMN and box C/D snoRNP biogenesis

The SMN complex was linked to box C/D snoRNP biogenesis in a yeast two-hybrid screen that indicated that the SMN protein interacts with the box C/D snoRNP common core protein fibrillarin (Liu and Dreyfuss, 1996; Jones et al, 2001). This interaction was confirmed by *in vitro* interaction assays and immunoprecipitation assays using human cell extracts (Jones et al, 2001; Pellizzoni et al, 2001). Mutational analysis of fibrillarin revealed that the N terminal domain RGG motif of fibrillarin mediates the interaction with SMN (Pellizzoni et al, 2001; Jones et al, 2001). The identity of the SMN domain that interacts with fibrillarin, however, is not clear with one study indicating the YG box (Pellizzoni et al, 2001) while another study proposing the tudor domain (Jones et al, 2001).

SMN is also essential for the correct localisation of box C/D snoRNP common core proteins and snoRNA. For instance, in human cells depleted of SMN both fibrillarin and NOP56, which are normally found in Cajal bodies and the nucleolus, were displaced (Lemm et al, 2006). Furthermore, the expression of a dominant negative SMN (SMN Δ N27), deficient in snRNP assembly, resulted in the redistribution of fibrillarin and U3 snoRNA to structures at the nucleoli periphery, which also contain SMN Δ N27 (Pellizzoni et al, 2001). Taken together this reveals that the presence of functional SMN is vital for the localisation of snoRNP factors and suggests that SMN is involved in box C/D snoRNP biogenesis.

SMN also interacts with the box H/ACA core protein GAR1 and the expression of SMN Δ N27 results in the mislocalisation of GAR1 (Jones et al, 2001; Pellizzoni et al, 2001) indicating that SMN functions in the biogenesis of both box C/D and H/ACA snoRNPs.

1.8. Sub nuclear structures

The nucleus is a highly organised structure containing a variety of non membrane bound compartments and nuclear bodies (Figure 1.19). A number of these nuclear compartments and bodies have been implicated in the biogenesis and function of snoRNPs, scaRNPs and the snRNPs and these include the nucleolus, the Cajal bodies, nuclear Gems and speckles.

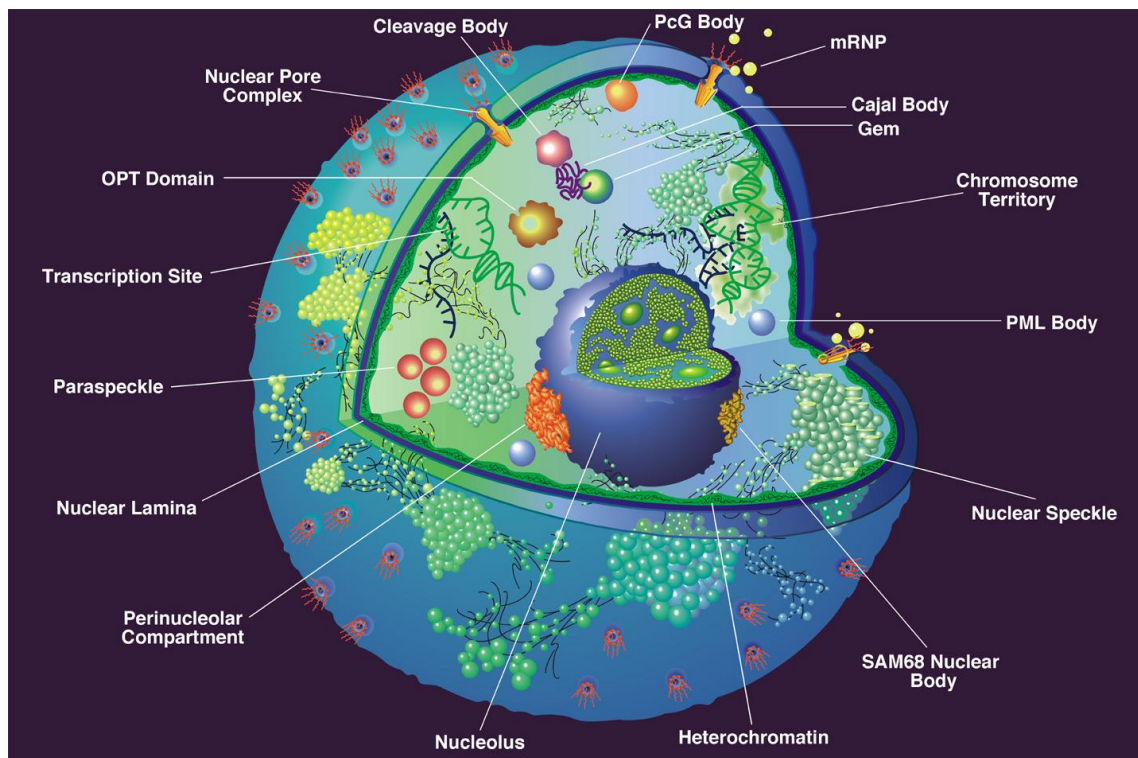


Figure 1.19: Sub nuclear structures

Schematic representation of the mammalian nucleus highlighting the various sub-nuclear bodies. Image taken from Spector, 2001.

1.8.1. The nucleolus

The most prominent feature of the nucleus is the nucleolus, which is between 3-5 μm in diameter (Sirri et al, 2008), and the site where the majority of ribosome synthesis occurs. The nucleolus is formed around clusters of rDNA genes arranged in tandem repeats known as NORs. In mammalian cells there may be numerous NORs, which results in the formation of multiple nucleoli. The nucleolus consists of three sub-

compartments, the inner FC, the DFC and the GC (Figure 1.20; Schwarzacher and Mosgoeller, 2000). The rDNA genes are located, either within, or at the periphery of the FC and are transcribed by RNA polymerase I at the border between the FC and DFC (Chooi and Leiby, 1981). As the pre-rRNA transcript migrates from the FC through the DFC and finally the GC it is modified and processed by factors that are localised to these compartments.

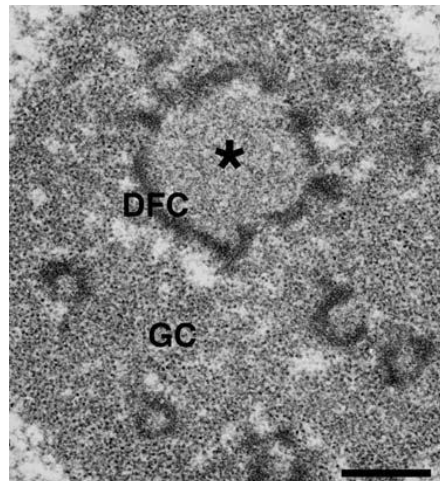


Figure 1.20: The nucleolus

An electron microscope image of a human nucleolus. The HeLa cell fixed in glutaraldehyde and osmium tetroxide, embedded in Epon and the section contrasted with uranyl acetate and lead citrate. The FC is shown by asterisk and is bordered by the DFC which is immersed in the GC. Multiple smaller nucleoli are present in addition to the indicated nucleoli. Bar embedded bottom right of image represents 0.5 μm . Image taken from Sirri et al, 2008.

The nucleolus is a dynamic structure, which is disassembled during mitosis and reassembled during G1 phase (Leung et al, 2004). Nucleoli also respond to changes in metabolic activity and growth rate in order to produce a sufficient supply of ribosomes. Given the central role of the nucleolus it is not surprising that it is a tightly regulated and has also been implicated in cell cycle regulation and stress responses (Visintin and Amon, 2000).

Proteomic analysis revealed that the human nucleolus contains over seven hundred proteins (Scherl et al, 2002; Andersen et al, 2002, 2005) with only one third of these known to be involved in ribosome biogenesis (Andersen et al, 2005). This could indicate that processes other than ribosome biogenesis occur in the nucleolus and based

on the identified proteins these may include cell cycle regulation, DNA repair and mRNA processing (Boisvert et al, 2007).

1.8.2. Cajal bodies

Cajal bodies (previously known as coiled bodies) are of around 0.1-2.0 μm in diameter and depending on cell type and the stage of the cell cycle there are between one and ten per cell (Figure 1.21; Spector, 2006). These bodies have been proposed as sites of box C/D snoRNP biogenesis (Verheggen et al, 2001, 2002; Boulon et al, 2004), snRNP modification (Jady and Kiss et al, 2001; Sleeman and Lamond, 1999), snRNP core protein assembly (Nesic et al, 2004) and U4/U6 di-snRNP and U4/U6.U5 tri-snRNP formation (Stanek et al, 2003; Schaffert et al, 2004). The protein coilin, which is vital for the formation of Cajal bodies, is often used as a marker of the Cajal bodies and most likely functions as a structural component (Figure 1.21; Kaiser et al, 2008). SMN is also enriched in Cajal bodies and has not only been shown to interact with coilin but is also vital for Cajal body formation (Liu and Dreyfuss, 1996; Matera and Frey, 1998; Shpargel and Matera, 2005; Lemm et al, 2006; Kaiser et al, 2008). The role of SMN in the Cajal bodies is not clear; however, the import of SMN to Cajal bodies has been shown to be dependent upon snRNP biogenesis (Matera and Shpargel, 2005).

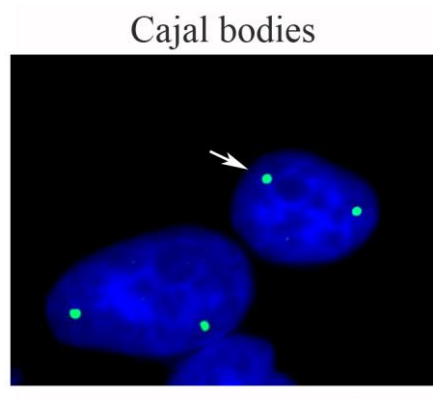


Figure 1.21: Cajal bodies

HeLa SS6 cells were grown on coverslips and fixed using paraformaldehyde. To detect Cajal bodies immunofluorescence was performed using anti-coilin antibodies (green). DAPI staining of DNA was used to indicate the nucleoplasm (blue). White arrow indicates a Cajal body. The bar to the bottom right of the panel represents 5 μm .

1.8.3. Nuclear Gems

Nuclear Gems are highly enriched in SMN and are sometimes associated with Cajal bodies. These bodies appear to be distinct from Cajal bodies, as they do not contain the Cajal body marker coilin or snRNPs. The function of this body, however, is not clear which has led to the proposal that the Gems and Cajal bodies may simply be different manifestations of the same structure (Carvalho et al, 1999; Young et al, 2001; Herbert et al, 2002).

1.8.4. Nuclear speckles

Pre-mRNA splicing factors, including the snRNPs, are distributed throughout the nucleoplasm and enriched in nuclear speckles, of which there are between twenty-five and fifty per cell (Spector, 1993; Fu, 1995). Larger nuclear speckles often correspond to interchromatin granule clusters (IGCs) and are irregular in shape and around 0.8-1.8 μm in diameter (Spector, 2006). Nuclear speckles are very dynamic and appear to be storage sites for pre-mRNA splicing factors rather than active sites of splicing (Misteli et al, 1997).

1.8.5. OPT (oct1 / PTF / transcription) domains

Transcription sites are diffusely located in the nucleoplasm; however, a number of transcription factors are concentrated in OPT (oct1 / PTF / transcription) domains. There are between one and three OPT domains per cell which are often found located close to the nucleolus being between 1.0-1.5 μm in diameter. The OPT domains consists of transcription factors and nascent transcripts; however, their function is not clear (Pombo et al, 1998).

1.8.6. Cleavage bodies

In addition to being diffusely localised in the nucleoplasm a number of factors involved in mRNA 3' processing are concentrated in cleavage bodies. There are one and four

cleavage bodies per cell that are between 0.3-1.0 μm in diameter, which are sometimes localised nearby Cajal bodies (Schul et al, 1996).

1.8.7. Perinucleolar compartment

The perinucleolar compartment is found associated with the surface of the nucleoli and has been implicated in RNA metabolism. There are between one and four perinuclear compartments per cell of 0.3-1.0 μm in diameter and contain RNA polymerase III transcripts as well as RNA binding proteins (Huang, 2000). The function of the perinucleolar compartment is not clear but interestingly these structures are predominately found in cancer cells (Huang, 2000).

1.8.8. PML (promyelocytic leukaemia) bodies

PML bodies have been linked to transcriptional regulation and nuclear protein sequestration. There are between ten and thirty PML bodies per cell that are 0.3-1.0 μm in diameter (Maul et al, 2000).

1.9. Research aims

The biogenesis of box C/D snoRNPs in human cells has been shown to occur in large, dynamic multiprotein pre-snoRNP complexes that consist of snoRNA, common core proteins and a variety of assembly, transport and RNA processing factors (Boulon et al, 2004; Watkins et al, 2004; 2007; McKeegan et al, 2007). The precise function of many of these factors and the processes required for the association of the common core proteins with the box C/D snoRNA have yet to be fully defined. Furthermore, the location of box C/D snoRNP biogenesis is not clear with data indicating the nucleoplasm (Terns et al, 1995; Watkins et al, 2004; Boulon et al, 2004), Cajal bodies (Verheggen et al, 2001; Boulon et al, 2004; Lemm et al, 2006) and cytoplasm (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007).

As the SMN complex has been linked to box C/D snoRNP biogenesis through an interaction with fibrillarin this study set out to further analyse the association of fibrillarin with the box C/D snoRNP and the role of the SMN complex in this process. These aims were broken down into the following objectives:

- Develop assays to investigate box C/D snoRNP biogenesis
- Investigate the association of fibrillarin with the box C/D snoRNP
- Analyse the interactions of fibrillarin with box C/D snoRNP assembly factors
- Characterise the interactions between the SMN complex and box C/D pre-snoRNP factors
- Determine the role of the SMN complex in box C/D snoRNP biogenesis
- Analyse the interaction of Snurportin1 with the box C/D pre-snoRNP and the SMN complex

Chapter two

Materials and methods

2.1. General molecular biology

2.1.1. Polymerase chain reaction (PCR)

PCR (polymerase chain reaction) was used to amplify DNA (deoxy ribonucleic acid) of interest. PCR reactions were set up using 100-500 ng (nanograms) of template DNA, 1 x Qiagen PCR buffer (containing 15 mM (milliMolar) MgCl_2), 0.2 mM dNTP (deoxyribonucleotide triphosphate), 200 mM forward primer, 200 mM reverse primer and 2.5 units (U) of HotstarTaq DNA polymerase (Qiagen), in a total volume of 100 μl (microlitre). The HotstarTaq DNA polymerase required incubation at 95 °C for 15 minutes for activation.

2.1.2. Agarose gel electrophoresis

1 x loading buffer (40 % glycerol (v/v), 60 % TE buffer (Tris-ethylene diamine tetra-acetic acid [EDTA]) (v/v), 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and 1 x bromophenol blue was added to the DNA to be analysed. DNA samples were loaded onto an appropriate percentage agarose gel, containing 0.035 $\mu\text{g}/\text{ml}$ (microgram per millilitre) ethidium bromide, and run in 1 x TAE (Tris-Acetate EDTA) buffer (400 mM Tris-HCl, 20 mM acetic acid, 0.1 mM EDTA, pH 8.0) at 80 volts for 1 hour. The gel was visualised using ultraviolet light and a Syngene gel imaging and documentation system.

2.1.3. DNA extraction and purification from agarose gel

DNA was removed from the agarose gel by means of a razor blade and purified using a QIAquick gel extraction kit (Qiagen), according to manufacturers guidelines.

2.1.4. DNA quantification

DNA concentration was determined by spectrophotometry with absorbance measured at wavelengths of 260 and 280 nm (nanometers). The purity of DNA was analysed by the ratio of absorbance at 260 / 280 nm.

2.1.5. Restriction enzyme digestion

Typically 0.2-1.5 µg of DNA was digested using 10 U of the required restriction endonuclease (Promega) in a final volume of 20 µl, which also contained 1 x restriction enzyme buffer (as described by manufacturer). Digestion was performed at 37 °C for two hours.

2.1.6. Ligation of isolated DNA into plasmid

Ligation reactions were performed with a three to one ratio of insert DNA to plasmid DNA. A typical ligation reaction contained 3 U of T4 DNA ligase (Promega), 1 x T4 DNA ligase buffer (as described by manufacturer) and DNA in a final volume of 10 µl, which was incubated overnight at 4 °C.

2.1.7. Transformation of bacteria with plasmids

To propagate plasmids 50 ng of DNA (or 5 µl of a ligation reaction (Chapter 2.1.6)) was added to 100 µl of chemically competent DH5α *Escherichia coli* (*E. coli*; provided by K.S.McKeegan, Newcastle University, UK) and cells stored on ice for 20 minutes. The cells were subjected to heat shock treatment at 42 °C for 45 seconds and incubated on ice. After 2 minutes 200 µl of luria broth (LB) was added to the cells and the sample incubated for 1 hour at 37 °C with constant agitation. Following incubation the cells were streaked on LB agar plates, containing appropriate antibiotics for the selection of successful transformations, and plates incubated at 37 °C overnight. Single colonies were picked from the LB plates and grown in 5 mls of LB, containing the appropriate antibiotic for selection. The samples were grown overnight at 37 °C in a rotating incubator set at 225 revolutions per minute (rpm). DNA was extracted from the samples using a QIAprep Spin Miniprep kit (Qiagen), according to manufacturers guidelines.

2.1.8. DNA sequencing

To confirm that the correct insert had been cloned into the plasmids restriction digests were performed as described in Chapter 2.1.5. Furthermore, to ensure that no mutations

had occurred during PCR plasmids were sent for DNA sequencing, which was performed by GATC Biotech.

2.2. DNA constructs

2.2.1. U3 and U14 box C/D snoRNA pcDNA5 expression plasmids

To create a pcDNA5/FRT/TO plasmid that expressed mouse U14 (mU14) snoRNA a section of the mouse hsc70 gene, consisting of intron 5 (containing the U14 snoRNA) and sections of the surrounding exons 5 and 6, was amplified by PCR (primers listed in Table 2.1) from an hsc70 expression plasmid (pRLMU14.5; Watkins et al, 1996). The amplified DNA was cloned into the *Bam*HI and *Xho*I restriction sites of the pcDNA5/FRT/TO plasmid (Invitrogen).

| Primer | Sequence |
|----------------------|--|
| mU14 (hsc70) forward | 5'-GCGCGGTACCCTGCAAGACTTCTTCAATGG-3' |
| mU14 (hsc70) reverse | 3'-GCGCCTCGAGGAGTGACATCCAAGAGC-5' |
| StreptoTag U3 A-1 F | 3'-CTGATAGGGATTCTCTATCACTGATAGGGACAAACTTTTAACTGATGGCG-5' |
| StreptoTag U3 A-1 R | 5'-GTGATAGAGAATCCCTATCAGTGATAGAGATTAAGACTATACTTTCAGGG-3' |
| StreptoTag U3 A-2 F | 3'-GATAGGGATGCTATCTCTATCACTGATAGGGACTTTTAACTGATGGCGAGAAAC-5' |
| StreptoTag U3 A-2 R | 5'-GATAGAGATAGCATCCCTATCAGTGATAGAGATTAAGACTATACTTTCAGGG-3' |
| StreptoTag U3 B F | 5'-ACGCGTATGCGAGATCTCTGCAGCTTTGAA-3' |
| StreptoTag U3 B R | 3'-ACGCGTATGCGAGATCTCCCGAGATCG-5' |

Table 2.1: Primers used to create mU14 and StreptoTag U3 pcDNA5/FRT/TO expression plasmids

To create a pcDNA5/FRT/TO plasmid that expressed human U3 snoRNA under Tet regulation a Tet-operator 2 (TetO₂) sequence (5'-TCCCTATCAGTGATAGAGA-3') was inserted into the U3 snoRNA promoter region. To create these Tet inducible U3 snoRNAs PCR was performed on a pCR4 TOPO plasmid (Invitrogen) that expressed human U3 snoRNA with a streptavidin tag (StreptoTag U3; Granneman et al, 2004). The first round of PCR was performed using either the StreptoTag U3 A-1 or A-2 primers (Table 2.1), which inserted the TetO₂ sequence between the PSE and U3 snoRNA sequence (Figure 2.1). The use of different primer combinations inserted the TetO₂ sequence in slightly different locations (Figure 2.1). The product of the first round

of PCR was then amplified using the StreptoTag U3 B primers (Table 2.1) and cloned into the *mlu1* restriction site of the pcDNA5/FRT/TO expression plasmid, upstream of the CMV (cytomegalo virus) promoter (see Figure 2.2).

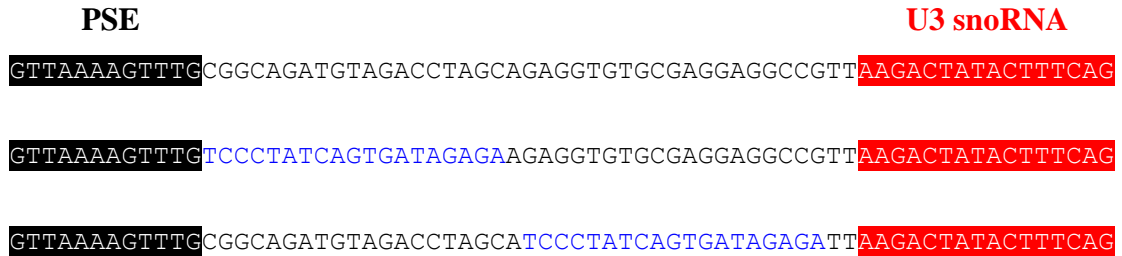


Figure 2.1: Tet-responsive U3 box C/D snoRNA promoter

The top panel shows a section of the U3 box C/D snoRNA gene with the PSE (proximal sequence element; white lettering on black background), U3 snoRNA sequence (white lettering on red background) and intervening sequences (black lettering). Middle and bottom panels show the location of the inserted Tet⁰₂ operator sequences (Blue text).

2.2.2. Fibrillarin and SMN pcDNA5 expression plasmids

Human fibrillarin and SMN were both cloned into a modified a pcDNA5/FRT/TO expression plasmid, which contained DNA coding for two FLAG tags (N-DYKDDDDK-C; 2 x FLAG) and a tag consisting of six histidines (6 x His), both downstream of the transcription start site (modified plasmid provided by A.Knox, Newcastle University, UK). These tags were expressed fused to the N terminal domain of any insert cloned into the multiple cloning site (Figure 2.2). The DNA inserted into the pcDN5/FRT/TO expression plasmid, which codes for the 2 x FLAG and 6 x His tags, is as follows:

5'-ATG GAC TAC AAA GAC GAT GAC GAT AAA GAC TAC AAA GAC GAT GAC GAT AAA GGT CAC CCA GGA TCA CTG GAA GTT CTG TTC CAG GGG CCC CTG CAT CAC CAC CAT CAC CAT GGA TC-3'

Fibrillarin cDNA was amplified by PCR using primers in Table 2.2 from a pCINeo VSV-fibrillarin expression plasmid (provided by G.Pruijn, Nijmegen centre for molecular life sciences, Netherlands). SMN cDNA was amplified using primers from Table 2.2 from a SMN pET17b expression plasmid (provided by U.Fischer, University

of Wurtzberg, Germany). The fibrillarin and SMN cDNAs were cloned into *Bam*HI and *Xho*I restriction sites of the modified pcDNA5/FRT/TO expression plasmid.

| Primer | Sequence |
|---------------------|--|
| fibrillarin forward | 5'- CGCGGGATCCATGAAGCCAGGATTCAGTCCCCG-3' |
| fibrillarin reverse | 3'-CGCGCTCGAGTCAGTTCCTCACCTTGGGGGGTGG-5' |
| SMN forward | 5'-CGGGGGATCCATGGCGATGAGCAGCGGCGG-3' |
| SMN reverse | 3'-CGCCTCGAGTTAATTTAAGGAATGTGAGCACC-5' |

Table 2.2: Primers used to create fibrillarin and SMN pcDNA5/FRT/TO expression plasmids

2.2.3. Degradation domain (DD) proteotuner plasmid

To express human fibrillarin fused to an N terminal degradation domain (DD) fibrillarin cDNA with 2 x FLAG and 6 x His tags was amplified by PCR (Table 2.3), from the modified pcDNA5/FRT/TO plasmid (Chapter 2.2.2.), and cloned into the *eco*R47III and *Xho*I restriction sites of the pPTuner IRES2 reporter plasmid (Clontech; DD-fibrillarin reporter plasmid).

| Primer | Sequence |
|--------------------|--|
| pTuner fib forward | 5'-GCGCAGCGCTATGGACTACAAAGAC-3' |
| pTuner fib reverse | 3'-CGCGCTCGAGTCAGTTCCTCACCTTGGGGGGTGG-5' |

Table 2.3: Primers used to create DD-fibrillarin reporter plasmid

2.2.4. Fibrillarin deletion mutants

Human fibrillarin deletion mutants were produced by PCR amplification (Table 2.4) of human fibrillarin cDNA from a pCINeo VSV-fibrillarin plasmid (provided by G.Pruijn, Nijmegen centre for molecular life sciences, Netherlands). The amplified fibrillarin deletion mutants (with the exception of N-fibrillarin) were cloned into pGEMT-easy plasmids (Promega) as described by the manufacturer. The N-fibrillarin deletion mutant

was amplified by PCR and cloned into a pET100 directional plasmid (Invitrogen) as described by the manufacturer. The fibrillarin deletion mutants were sub cloned from the pGEMT-easy plasmids into pECFP-C1 plasmids (Clontech) using *Hind*III and *Bam*H1 restriction sites and also into pET17b expression plasmids (Invitrogen) using *Hind*III and *Xba*I restriction sites.

| Primer | Sequence |
|---|--|
| full length fibrillarin | 5'-ATGAAGCCAGGATTCAGTCCCCG-3' |
| | 3'-TCAGTTCTTCACCTTGGGGGGTGG-5' |
| R-fibrillarin (amino acids 1-81) | 5'-ATGAAGCCAGGATTCAGTCCCCG-3' |
| | 3'-TCACGACTGGTTTCCTCTTTTCCTCCC-5' |
| R-N-fibrillarin (amino acids 1-141) | 5'-ATGAAGCCAGGATTCAGTCCCCG-3' |
| | 3'-TCAGGGGTTCCAGGCTCGGTACTC-5' |
| N-C-fibrillarin (amino acids 82-321) | 5'-ATGGGGAAGAATGTGATGGTGGAG-3' |
| | 3'-TCAGTTCTTCACCTTGGGGGGTGG-5' |
| C-fibrillarin (amino acids 142-321) | 5'-ATGCGCTCCAAGCTAGCAGCAGCAATCC-3' |
| | 3'-TCAGTTCTTCACCTTGGGGGGTGG-5' |
| N-fibrillarin (amino acids 82-141) | 5'-CACCGGATCCATGGGGAAGAATGTGATGGTGGAG-3' |
| | 3'-CTCGAGTCAGGGGTTCCAGGCTCGGTACTC-5' |

Table 2.4: Primers used to create fibrillarin deletion mutants

2.3. Cell culture, cell transfection and extract preparation

2.3.1. Cell culture

Human HeLa SS6 (cervical carcinoma) and human HEK293 (embryonic kidney) cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Sigma). Media was supplemented with 100 µg/ml streptomycin (Sigma), 100 U/ml penicillin (Sigma) and 10 % foetal bovine serum (FBS, Sigma). Cells were grown at 37 °C with 5 % carbon dioxide in a humidified incubator and split when 80 % confluent using trypsin (Sigma).

For long-term storage an 80 % confluent flask of cells was washed in phosphate buffered saline (PBS) and cells harvested using 1 ml of trypsin. The trypsin in the harvested cells was neutralised by the addition of 10 x volume of DMEM media and

cells pelleted using a swing bucket centrifuge at 600 rpm for 5 minutes. The supernatant was removed from the cell pellet and the cells resuspended in 1ml of freezing buffer (10 % DMSO, 90 % FBS) and dispensed into cryo-vials (Greiner). The cryo-vials containing cells were stored at -80°C overnight then transferred to liquid nitrogen.

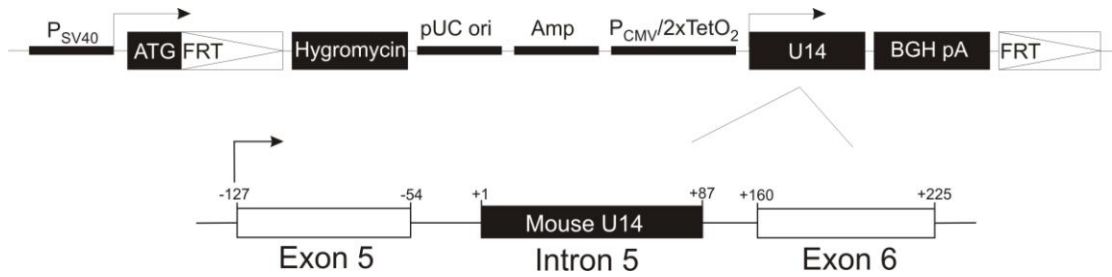
To prepare cells from a frozen stock the cryo-vials (containing frozen cells) were rapidly thawed in a 37°C water bath and transferred to a sterile 10 ml centrifuge tube (Greiner) and diluted in 10 mls of warm DMEM media. The cells were then centrifuged at 600 rpm for 5 minutes, supernatant discarded and pellet resuspended in 10 mls DMEM media and transferred to a cell culture flask.

2.3.2. Development of inducible HEK293 cells

The Tet-inducible mU14, U3 box C/D snoRNA, FLAG-fibrillarin, FLAG-SMN and control HEK293 cell lines were developed using the Flp-In T-Rex HEK293 cell system, as described by the manufacturer (Invitrogen).

The Flp-In T-Rex HEK293 cells contain a single integrated FRT (Flp Recombination Target) site and stably express the Tetracycline (Tet) repressor protein. To create inducible cell lines the gene of interest was cloned into a pcDNA5/FRT/TO expression plasmid, which contains a Tet-inducible promoter, FRT site and hygromycin resistance gene (see Chapter 2.2.1 and 2.2.2). The Flp-In T-Rex HEK293 cells were then co-transfected with the pcDNA5/FRT/TO plasmid (containing gene of interest) and the pOG44 plasmid (Invitrogen), which expresses Flp recombinase and mediates integration of pcDNA5/FRT/TO plasmid into FRT sites. Transformed HEK293 cells were grown in media supplemented with $10\text{ }\mu\text{g/ml}$ of blastocidin and $100\text{ }\mu\text{g/ml}$ of hygromycin to select and maintain the Tet-repressor and pcDNA5/FRT/TO inserts, respectively. Cells were induced by the addition of $1\text{ }\mu\text{g/ml}$ Tet.

A



B



C

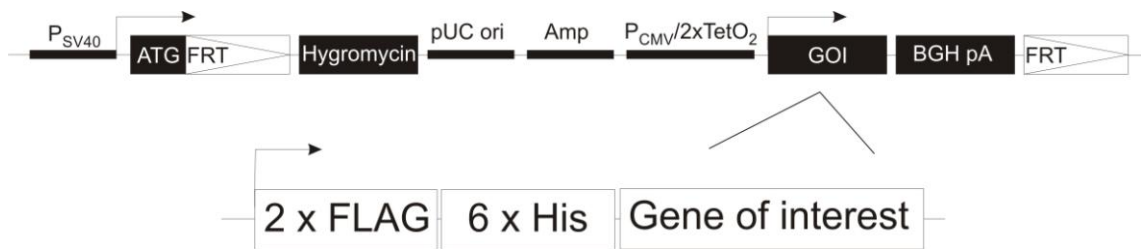


Figure 2.2: Schematic representation of integrated pcDNA/FRT/TO plasmids

Schematic diagram of Tet-inducible (A) mU14 (B) human U3 box C/D snoRNAs and (C) 2 x FLAG and 6 x His tagged SMN and fibrillarin in a pcDNA5/FRT/TO expression cassette incorporated into the Flp-In T-Rex HEK293 FRT site. The incorporation of the pcDNA5/FRT/TO plasmid at the FRT recombination site inserts a P_{SV40} (Simian virus 40 promoter) and ATG start site to the hygromycin resistance gene. The pUC origin indicated was required for pcDNA5/FRT/TO plasmid propagation in *E. coli* and the Amp (Ampicillin gene) was used for selection. BGH pA (Bovine growth hormone polyadenylation signal) was required for efficient termination and polyadenylation of mRNA expressed from the CMV promoter. (B) The StreptoTag U3 snoRNA gene was inserted upstream of the CMV promoter as it contained its own Tet-inducible promoter.

2.3.3. Electroporation

An Amaxa II electroporator was used to transfect HeLa SS6 and HEK293 cells with siRNA duplexes and plasmids. Optimal protocols were used for transfection of HeLa SS6 cells (programme I-013, Kit R; Amaxa) and HEK293 cells (programme A-023, Kit V; Amaxa) as described by manufacturer. For each million cells 2 µg of plasmid DNA or siRNA was transfected.

2.3.4. Chemical transfection of plasmids

Cells were grown in 0.5 ml tissue culture wells and transfected with 2 µg of plasmid DNA using FuGENE 6 transfection reagent (Roche Applied Science) as described by manufacturer.

2.3.5. siRNA mediated depletion

To deplete target proteins in human cells siRNA mediated depletion was utilised. As a control, a siRNA duplex targeting firefly luciferase mRNA (GL2), which is not present in either HeLa SS6 or HEK293 cells was used. The GL2 siRNA duplex has no effect on HeLa SS6 cell growth or RNA levels (Elbashir et al, 2002).

To anneal siRNAs 20 µM of each sense siRNA strand was added to annealing buffer (100 mM KOAc, 2 mM MgOAc, 30 mM HEPES-KOH (4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid) [pH 7.4]), denatured at 90 °C for 1 minute and then incubated at 37 °C for 1 hour (Elbashir et al, 2002). The siRNA duplexes were stored at –20 °C.

HeLa SS6 cells or HEK293 cells were transfected with siRNA duplexes using electroporation as described in Chapter 2.3.3. For chemical transfection 20,000 cells were grown on glass coverslips in 0.5 ml cell culture wells. Twenty-four hours later cells were transfected with 3 µl of 20 µM siRNA duplexes (per well; Table 2.5) using oligofectamine (Invitrogen) as described by Elbashir et al, 2002 and cells harvested or fixed after 60 hours.

| Target | Accession | siRNA sense sequence | Source |
|-------------|-----------|-------------------------------|---------------------------------------|
| GL2 | X65324 | 5'-CGUACGCGGAUACUUCGAUU-3' | Elbashir et al, 2002 |
| NOP56 | NM_006392 | 5'-CAAUAUGAUAUCCAGUCCAUA-3' | Watkins et al, 2004 |
| NOP58 | NM_015934 | 5'-CAAGCAUGCAGCUUCUACCGUUC-3' | Watkins et al, 2004 |
| fibrillarin | NM_001436 | 5'-CAGUCGAGUUCUCCACCGCUCU-3' | Watkins et al, 2004 |
| SMN | XM_041492 | 5'-GUGGAAUGGGUAAUCUUCUU-3' | I.Lemm, Max Planck Institute, Germany |
| Gemin2 | NM_003616 | 5'-GUACAGAUUUAACACAUCUU-3' | I.Lemm, Max Planck Institute, Germany |
| Gemin5 | BI868900 | 5'-GUUUGAAUCAACCCUGCCUUU-3' | I.Lemm, Max Planck Institute, Germany |
| Gemin6 | NM_024775 | 5'-GACUAUAUUUAUCCCUCAUU-3' | I.Lemm, Max Planck Institute, Germany |
| Gemin7 | NM_024707 | 5'-CCUUCAAGCCAUAAGAUAAdTdT-3' | I.Lemm, Max Planck Institute, Germany |
| hPRP31 | NM_015629 | 5'-AAGCCAAAGCUUCAGAAGUGAUG-3' | Schaffert et al, 2004 |

Table 2.5: The siRNA duplexes

2.3.6. Cell extract preparation

For extract production two 162 cm² flasks of cells were grown to 80 % confluency and harvested using trypsin. Cells were pelleted by centrifugation in an Eppendorf minifuge set at 300 rpm for 1 minute and washed with PBS.

To produce whole cell extracts the cell pellet was resuspended in 1 ml of cold GL50E buffer (1.5 M KCl, 200 mM HEPES, 5 mM EDTA) and sonicated (on ice) using a Bandelin Sonopuls HD2070 ultrasonic homogeniser (with a 2 mm diameter MS72 titanium microtip) set at (6 x) 20 seconds continuous sonication (at 30 % power) with 40 seconds off (in between pulses). Triton-X100 was then added to the sample to a concentration of 0.2 % (v/v) and insoluble material removed by centrifugation (using a Eppendorf minifuge) at 13,000 rpm (4 °C) for 10 minutes (Granneman et al, 2004). The supernatant was removed and pellet discarded. The cell extract was then either frozen in liquid nitrogen and stored at –80 °C or used directly in an assay.

To produce nucleoplasmic and nucleolar extracts cells were harvested as described for whole cell extracts. The following protocols were adapted from Jones et al, 2001 and all steps were performed on ice or in centrifuges at 4 °C. Cell pellets were resuspended in 0.5 mls of cold lysis buffer 1 (10 mM Tris-HCl [pH 7.5], 10 mM KCl, 2 mM MgCl₂, 0.05 % Triton-X100 (v/v), 1 mM DTT (dithiothreitol) / 1 x complete mini protease

inhibitor cocktail from Roche) and swelled on ice for 20 minutes. Crude nuclear extract was then produced by forcing the cell lysate through a 25 gauge needle (5 x) followed by centrifugation (using a Eppendorf minifuge) at 3,000 rpm for 5 minutes. The nuclear pellet was resuspended in 0.5 ml lysis buffer 1 and sonicated using a Bandelin Sonopuls HD2070 ultrasonic homogeniser (with a 2 mm diameter MS72 titanium microtip) set at (6 x) 20 seconds continuous sonication (at 30 % power) with 40 seconds off (in between pulses). The nucleoli and nucleoplasm was separated by centrifugation (using a Eppendorf minifuge) at 8,000 rpm for 5 minutes. The supernatant contained nucleoplasmic extract while the pellet was enriched in nucleoli. The nucleoplasmic extract was then either snap frozen in liquid nitrogen and stored at -80 °C or used directly in an assay.

The enriched nucleoli pellet was resuspended in 0.5 ml lysis buffer 2 (10 mM Tris-HCl [pH 7.5], 300 mM KCl, 2 mM MgCl₂, 0.05 % Triton-X100 (v/v), 1 mM DTT, 1 x complete mini protease inhibitor cocktail from Roche) and swelled on ice for 45 minutes. The sample was then subjected to sonication, as performed with the nucleoplasmic extract, and insoluble material removed by centrifugation (using an Eppendorf minifuge) at 13,000 rpm for 5 minutes. The supernatant was removed (containing the nucleoli extract) and either snap frozen in liquid nitrogen and stored at -80 °C or used directly in an assay.

2.3.7. Glycerol gradients preparation

To produce a 10-30 % glycerol gradient 6.5 mls of a 10 % glycerol solution (150 mM KCl, 20 mM HEPES, 1.5 mM MgCl₂, 10 % glycerol (v/v), 0.02 % Triton-X100 (v/v), 1 mM DTT) was added to a Beckman SW40 gradient tube. A 30 % glycerol solution (150 mM KCl, 20 mM HEPES, 1.5 mM MgCl₂, 30 % glycerol, 0.2 mM DTT) was then added to the bottom of the tube using a large needle until the interface between the two solutions reached halfway up the tube. The 10-30 % glycerol gradient was made on a gradient master 107ip (Biocomp) set to the SW40 protocol at an angle of 81.5, speed 15 for 2 minutes 20 seconds. The glycerol gradients were then chilled at 4 °C for 1 hour.

Cell extract was applied to the top of the 10-30 % glycerol gradient and the sample centrifuged in Beckman optima L-100 ultra centrifuge using a Sw60ti rotor at a speed of 40,000 rpm for 3 hours 30 minutes at 4 °C with acceleration and deceleration set at 5. Fractions were collected from the top of the gradient.

2.3.8. Immunoprecipitation (IP)

Protein G Sepharose beads (GE Healthcare) were washed (3 x) with 1 ml PBS containing 0.1 % Triton-X100 (v/v), centrifuged at 300 rpm using an Eppendorf minifuge for 1 minute and supernatant discarded. Antibodies against the protein of interest were added to 10 µl of washed protein G Sepharose beads, volume made up to 300 µl with PBS containing 1 % Triton-X100, and incubated on a head over tail blood wheel (at 6 rpm) for 1 hour at room temperature or overnight at 4 °C. Following incubation the protein G Sepharose beads were washed (3 x) using wash buffer (20 mM HEPES [pH 8], 150 mM NaCl, 3 mM MgCl₂, 0.5 mM DTT, 10 % glycerol (v/v), 0.1 % Triton-X100 (v/v)) as described earlier. Protein G Sepharose beads (with immobilised antibodies) were incubated with the cell extract of interest made up to 300 µl with wash buffer on a head over tail blood wheel (6 rpm) for 2 hours at 4 °C. As an input / control 10 % of cell extract was retained. IP samples were washed (3 x) with 1 ml of wash buffer (as above), transferred to a new tube (to reduce background) and precipitated factors removed from beads using homogenisation buffer (1 % SDS (w/v), 50 mM Tris-HCl [pH 7.5], 50 mM NaCl, 0.5 mM EDTA).

2.4. Protein separation and analysis

2.4.1. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were combined with 1 x protein loading buffer (75 mM Tris-HCl [pH 6.8], 1.25 mM EDTA, 20 % glycerol (v/v), 2.5 % SDS (w/v), 0.125 mM bromophenol blue), supplemented with 200 mM DTT and denatured at 95 °C for 5 minutes. Samples were loaded onto a denaturing SDS-PAGE consisting of a 4 % acrylamide stacking gel (4 % acrylamide (37.5:1 acrylamide: bisacrylamide), 125 mM Tris-HCl [pH 6.8], 0.1 %

SDS (w/v)) and a 12 % acrylamide resolving gel (12 % acrylamide (37.5:1 acrylamide: bisacrylamide), 375 mM Tris-HCl [pH 8.8], 0.1 % SDS (w/v)). The 12 % denaturing SDS-PAGE gel was run at 200 volts in 1 x running buffer (25 mM Tris [pH 8.3] / 250 mM glycine, 0.1 % SDS (w/v)) until dye front reached the end of the gel.

2.4.2. Coomassie blue staining

To visualise proteins the SDS-PAGE gel was stained using Coomassie blue staining solution (0.1 % Coomassie blue (w/v), 40 % methanol (v/v), 10 % acetic acid (v/v)) for 30 minutes with gentle agitation. The gel was then destained using destaining solution (40 % methanol (v/v), 10 % acetic acid (v/v)), which was replenished several times until the gel is fully destained, and proteins visible.

2.4.3. Western blot assays

Proteins separated on an SDS-PAGE gel were transferred to nitrocellulose membrane (Protran) using a Western blot transfer tank containing transfer buffer (25 mM Tris base, 150 mM glycine, 10% methanol, [pH 8.3]) run at 40 volts for 2 hours. Nitrocellulose membranes were incubated with blocking buffer (PBS, 3 % milk powder (w/v), 0.05 % tween 20 (v/v)) for 1 hour at room temperature or overnight at 4 °C. Antibodies were diluted to desired concentration in blocking buffer and incubated with the membrane for 2 hours at room temperature. Membranes were washed (3 x) for 5 minutes and (2 x) 15 minutes in PBS containing 0.05 % tween 20 (v/v). Secondary HRP (horse radish peroxidase) conjugate antibody was diluted in blocking buffer and incubated with the membrane for 2 hours and washed the same as with the primary antibody. Results were visualised using ECL detection (enhanced chemiluminescence; GE healthcare) and hyperfilm (Amersham) as described by the manufacturer.

| Antibody | Raised in | Raised against | Manufacturer / source | Code |
|---|-----------|------------------------|---|----------|
| anti-fibrillarin | rabbit | human fibrillarin | Santa Cruz | sc-25397 |
| anti-NOP56 | rabbit | human NOP56 peptide | N.J.Watkins; Watkins et al, 2000, 2002 | N/A |
| anti-NOP58 | rabbit | human NOP58 peptide | N.J.Watkins; Watkins et al, 2000, 2002 | N/A |
| anti-nucleolin | rabbit | human nucleolin | Abcam | ab22758 |
| anti-SMN (7B10) | mouse | human SMN | U.Fischer; Jones et al, 2001 | N/A |
| anti-Gemin2 (H-100) | rabbit | human Gemin2 | Santa Cruz | sc50404 |
| anti-FLAG M2 | mouse | FLAG sequence | Sigma | F3165 |
| anti-Poly histidine | mouse | polyhistidine sequence | Sigma | H1029 |
| anti-mouse immunoglobulin HRP conjugate | rabbit | mouse immunoglobulin | Dako | P0260 |
| anti-rabbit immunoglobulin HRP conjugate | donkey | rabbit immunoglobulin | Santa Cruz | sc-25397 |

Table 2.6: Antibodies used in Western blots assays

2.5. RNA extraction, separation and analysis

2.5.1. RNA extraction and separation

To extract RNA from a cell pellet or IP assay (of around 100 µl) 200 µl of homogenisation buffer (Chapter 2.3.8) and 200 µl of phenol: chloroform: isoAmyl Alcohol (25:24:1) [pH 6.6 / 8.0] was added, sample vortexed for ~30 seconds and centrifuged (using a Eppendorf minifuge) for 3 minutes at 13,000 rpm. The upper aqueous phase containing RNA was removed and added to a new tube. RNA was precipitated from the sample by the addition of 10 µl of 5 M NaCl and 1 ml of 100 % ethanol. The sample was then vortexed, stored at –80 °C for ~30 minutes and centrifugation performed (using a Eppendorf minifuge) for 15 minutes at 13,000 rpm. Supernatant was removed and RNA pellets left to dry. The resulting RNA pellet was dissolved in water and mixed 1:1 with RNA loading buffer (62.5 % formamide, 9.25 % formaldehyde, 1.25 x MOPS buffer (3-morpholinopropanesulfonic acid), 50 µg/ml bromophenol blue, 50 µg/ml xylene cyanol). Samples were denatured at 95 °C for 2 minutes then cooled on ice. RNA samples were separated by electrophoresis on an 8 %

acrylamide gel containing 7 M urea in 1 x TBE buffer and run at 350 volts until dye front reach the end of the gel.

2.5.2 Northern blot hybridisation

RNA that had been separated on an 8 % acrylamide gel containing 7 M urea (Chapter 2.4.1), was transferred to Hybond N membrane (Amersham biosciences) using a Trans-Blot ® Cell (BIO-RAD) overnight at 18 volts (4 °C) in transfer buffer (25 mM NaPO₄ [pH 6.4]). The RNA was then fixed to the membrane using an ultraviolet Stratalinker 2400 (Stratagene).

Radiolabelled probes against the U3, U8 and U14 box C/D snoRNAs and the U1 snRNA were produced as described by Watkins et al, 2004. To produce a radiolabelled probe against 7SL random primer labelling was utilised. The S domain of 7SL was amplified by PCR (using primers in Table 2.7) from a 7SL expression plasmid provided by K.Nagai, MRC laboratory of molecular biology, UK (Menichelli et al, 2007). Fifty nanograms of the 7SL PCR product was diluted in 10 µl H₂O, denatured at 95-100 °C for 5 minutes then placed on ice. Once cooled 3 µl of random hexamer mix (250 mM Tris [pH 7.5], 50 mM MgCl₂, 5 mM DTT, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 150 µg/ml random deoxribonucleotides from Amersham), 2 µl of α-³²P labelled dCTP and 5 U of klenow polymerase (Promega) was added, and the sample incubated at 37 °C for 30 minutes. To remove unincorporated nucleotides the 7SL probe was made up to 50 µl and spun through a G-50 spin column (GE Healthcare).

| Primer | Sequence |
|-------------|------------------------------|
| 7SL forward | 5'-CTATGCCGATCGGGTGTCCGC-3' |
| 7SL reverse | 3'-CAGCACGGGAGTTTTGACCTGC-5' |

Table 2.7: Primers used to create 7SL probe

Northern blot analysis was performed by incubating the Hybond N membrane, containing the fixed RNA, with hybridisation buffer (25 mM NaPO₄ [pH 6.5], 6 x SSC (Sodium chloride- sodium citrate), 5 x Denhardts, 0.5 % SDS (w/v), 50 % deionised formamide, 100 µg/ml denatured salmon sperm DNA) for 2 hours at 42 °C. The

radiolabelled probe was denatured at 95-100 °C for 5 minutes then added directly to hybridisation buffer containing the membrane and incubated overnight at 42 °C. The membrane was washed once with 2 x SSC containing 0.5 % SDS (w/v) to remove excess probe, washed twice with 2 x SSC containing 0.5 % SDS (w/v) for 5 minutes at room temperature (with constant shaking) and twice with 2 x SSC containing 0.1% SDS (w/v) for 5 minutes at room temperature (with constant shaking). The final wash was performed with 2 x SSC containing 0.1% SDS (w/v) for 30 minutes at 50 °C (with constant shaking). The results were viewed by autoradiography using either Kodak Biomax MR / MS (maximum resolution / maximum sensitivity) film or with a PhosphorImager screen using a typhoon scanner.

2.5.3 Primer extension assays

Primers complementary to the human U3 and mouse U14 snoRNAs (Table 2.8) were radiolabelled using γ P32 ATP. Ten pmoles of primer was mixed with 10 U of T4 polynucleotide kinase (PNK; Promega), 1 μ l of T4 PNK buffer (Promega), 5 μ l γ P32 ATP and H₂O up to 10 μ l, and incubated at 37 °C. To remove unincorporated nucleotides the probe was made up to 50 μ l and spun through a G-50 spin column.

| Target | Sequence |
|--------------------------|---------------------------------|
| human U3 box C/D snoRNA | 5'-ACCACTCAGACCGGTTCTCTCCC-3' |
| mouse U14 box C/D snoRNA | 5'-GGAAGGAA CTAGCCAACACAGCAC-3' |

Table 2.8: Oligonucleotides used in primer extension assays

RNA was extracted from the samples using phenol chloroform extraction method as described earlier in Chapter 2.5.1; however, after phenol extraction the aqueous phase was mixed with an equal amount of chloroform amyl alcohol (24 parts chloroform 1 part amyl alcohol), sample vortexed for ~30 s and centrifuged using a Eppendorf minifuge for 3 min at 13,000 rpm. The purpose of the additional phase was to prevent phenol carry over. The upper layer containing the RNA was then removed and RNA precipitated using ethanol as described in Chapter 2.5.1.

To perform the primer extension RNA was added to 0.5 µl 10 x HY buffer (0.5 M Tris-HCl [pH 8.4], 0.6 M NaCl, 0.1 M DTT), 0.5 µl radioactive primer, 1 µl H₂O and the sample denatured at 96 °C for 1 minute. The sample was cooled on ice and then 4.3 µl H₂O, 0.5 µl 10 x RT buffer (0.5 M Tris-HCl [pH 8.4], 0.2 mM dNTPs, 0.1 M MgCl₂, 0.6M NaCl, 0.1 M DTT) and 5 U AMV reverse transcriptase (Promega) was added and the sample incubated for 45 minutes at 43 °C. Once the reaction was complete RNA loading buffer was added. the sample denatured at 95 °C for 2 minutes then separated on an 8% acrylamide gel containing 7 M urea. The gel was fixed fixing solution (10 % methanol 10 % acetic acid) for 10 minutes and dried. The result was viewed using either using a PhosphorImager screen and a typhoon scanner or by MS / MR film (Kodak).

2.6. Microscopy

2.6.1. Fluorescent *in Situ* hybridisation (FISH)

Cells grown on glass coverslips were washed (3 x) in PBS and fixed using PBS containing 4 % paraformaldehyde (w/v) for 20 minutes. After fixation cells were washed (3 x) in 70 % ethanol and rehydrated in PBS containing 5 mM MgCl₂ for 10 minutes and then incubated in 15 % formamide, 2 x SSC, 10 mM NaPO₄ [pH 7] for 10 minutes at room temperature. For each coverslip 1 µl of each FISH probe (10 µg/µl; Table 2.9) was added to 1 µl tRNA (10 µg/µl), 1 µl of sonicated salmon sperm DNA (10 µg/µl), 5 µl of 2 x formamide buffer (60 % formamide, 40 mM NaPO₄ [pH 7.0]), 10 µl of FISH hybridisation buffer (dextrane sulphate, 4 x SSC, 0.4 % BSA (bovine serum albumin) (w/v)) in a total volume of 20 µl (Taneja and Singer, 1992). Fish solution was placed on parafilm, coverslips placed face down on solution and hybridisation performed in a humidified incubator at 37 °C for 4 hours.

After hybridisation, coverslips were washed (2 x) in 15 % formamide, 2 x SSC, 10 mM NaPO₄ [pH 7.0]) at 37 °C for 30 minutes. Coverslips were then washed (2 x) in x SSC containing 0.1 % Triton-X100 (v/v) at room temperature for 15 minutes followed by a wash (2 x) in 1 x SSC containing 0.1 % Triton-X100 (v/v) at room temperature for 15 minutes. For DAPI (Sigma-Aldrich) staining the final wash was performed with 0.1

µg/ml of DAPI. The coverslips were then immersed briefly (3 x) in H₂O and (3 x) in 70% ethanol then mounted onto glass microscope slides using 3.5 µl Moviol.

| Probe | Sequence | Source |
|--------------|--|--------------------------------------|
| U3 snoRNA | 5'-CGCTCTACACGTTTCAGAGAACTTCTCTAGTAAC ACACTATAGAAATGATCCC-3' | N.J.Watkins; McKeegan et al, 2007 |
| U8 snoRNA | 5'-CGTTCTAATCTGCCCTCCGGAAGGAGGAAACAGG AAACAGGTAAGGATTATCCCACCC-3' | N.J.Watkins; Watkins et al, 2007 |
| U2 snRNA | 5'-CTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTAT CTGTT-3' | N.J.Watkins; McKeegan et al, 2007 |
| U4 snRNA | 5'-TCACGGCGGGGTATTGGGAAAAGTTTTCAATTAG CAATAATCGCGCCT-3' | R.Luhrmann; Schaffert et al, 2004 |
| 7SL | 5'-TGAGTCCAGGAGTTCTGGGCTGTAGTGCCTATGCC GATCGGGT-3' | J.Brown, Newcastle University, UK |

Table 2.9: FISH probes

2.6.2. Immunofluorescence

Cells grown on coverslips were washed (3 x) in PBS and fixed using PBS containing 4 % paraformaldehyde (w/v) for 20 minutes. After fixation, cells were washed (3 x) in PBS and incubated in PBS containing 1.0 % Triton-X100 for 15 minutes at room temperature. Cells were washed (4 x) with PBS and then incubated in 10 % FCS (v/v), PBS [pH 7.4], 0.1 % Triton-X100 (v/v). After incubation for 1 hour at room temperature (or overnight at 4 °C) 50 µl of PBS with 10 % FCS (v/v), containing desired concentration of primary antibody (Table 2.10), was placed on top of the coverslips. After 2 hours the coverslips were rinsed (3 x) and washed (3 x) for 10 minutes with PBS. The secondary antibody was then applied as for the primary antibody and washed as described above. For DAPI staining the final wash was performed with 0.1 µg/ml of DAPI. Finally coverslips were immersed (3 x) in water and (3 x) in 70 % ethanol prior to being mounted onto glass slides using 3.5 µl Moviol.

| Antibody | Raised in | Raised against | Manufacturer | Code |
|---|-----------|-----------------------|--------------|---------|
| anti-FLAG M2 | mouse | FLAG sequence | Sigma | F3165 |
| anti-BMS1 | rabbit | human BMS1 peptide | Eurogenetc | N/A |
| anti-coilin | rabbit | human Coilin | Santa Cruz | sc32860 |
| anti-mouse Immunoglobulin alexa 555 conjugate | donkey | mouse immunoglobulin | Invitrogen | A31570 |
| anti-rabbit immunoglobulin alexa 488 conjugate | donkey | rabbit immunoglobulin | Invitrogen | A21208 |

Table 2.10: Antibodies used in immunofluorescence

2.6.3. Widefield microscopy

Cells subjected to FISH and immunofluorescence were visualised using an Axiovert 200M inverted microscope with a Plan-Apochromat, 100 x / 1.40 oil DIC, ∞ / 0.17 objective (Zeiss). The Zeiss filter sets used were as follows: 02 (DAPI), 20 (Cy3), 26 (Cy5), 38 High Efficiency (GFP). The microscope was provided by J.Brown, Newcastle University, UK.

2.7. Protein expression, purification and *in vitro* interaction assays

2.7.1. Recombinant protein expression and purification

To produce recombinant N terminal GST (Glutathione S-Transferase) tagged fibrillarin, N-C-fibrillarin (amino acids 82-321), Snurportin1, NOP17, NUFIP, TIP48 and TIP49 the respective cDNAs were cloned into pGEX-6P-1 / 2 expression plasmids (Table 2.11). To express GST tagged proteins, chemically competent BL21 *E. coli* (provided by K.S.McKeegan, Newcastle University, UK) were transformed with the expression plasmid (containing the cDNA of the protein of interest) by heat shock at 42 °C for 45 seconds. Successful transformations were selected using the appropriate antibiotic on LB agar plates (see Chapter 2.1.7). Transformed *E. coli* were grown in large LB cultures at 37 °C in a shaking incubator (set at 225 rpm) until an optical density of 0.35-4 was

reached (with the absorbance of the spectrophotometer set at A600). The expression of protein of interest was induced by addition of IPTG (isopropyl-beta-D-thiogalactopyranoside), to final concentration of 1 mM, and cells left overnight at 18 °C in a shaking incubator (set at 225 rpm). Cells were harvested in a swing bucket centrifuge at 4200 rpm for 30 minutes and resuspended in 50 mls of buffer A (20 mM Tris-HCl [pH 8.0], 300 mM KCl, 0.1 % tween 20 (v/v), 10 % glycerol (v/v), 1 mM Tris-hydroxy-phosphine). The cell suspension was sonicated, on ice, for 3 x 30 second bursts at full amplitude and cell debris removed by centrifugation at 24,500 rpm for 30 minutes using a Beckman A-20 rotor. For each litre of original culture 100 µl of washed glutathione Sepharose beads (GE Healthcare; washed in buffer A) were incubated with cell supernatant, under constant agitation for 2 hours at 4 °C. The glutathione Sepharose beads with bound GST proteins were captured using an elution column and washed with 50 mls of cold buffer A. GST tagged proteins were eluted from the glutathione Sepharose beads by incubation with 2.5 mls of cold buffer A supplemented with 50 mM reduced glutathione for 1 hour. The reduced glutathione was removed from the protein sample using desalt columns (Amersham) and AKTA protein purification system. Purified proteins were then stored at –80 °C after snap freezing in liquid nitrogen.

| Protein (all human) | Plasmid | Source | References |
|--------------------------------------|-----------|--------------|--------------------------------|
| fibrillarin | pGEX-6P-1 | K.S.McKeegan | McKeegan et al, 2007 |
| N-C-fibrillarin (amino acids 82-321) | pGEX-6P-1 | K.S.McKeegan | McKeegan et al, 2007 |
| Snurportin1 | pGEX-6P-1 | K.S.McKeegan | McKeegan et al, 2009 (in prep) |
| NOP17 | pGEX-6P-1 | K.S.McKeegan | McKeegan et al, 2007 |
| NUFIP | pGEX-6P-1 | P.Cabart | Cabart et al, 2004 |
| TIP48 | pGEX-6P-1 | N.J.Watkins | McKeegan et al, 2007 |
| TIP49 | pGEX-6P-1 | N.J.Watkins | McKeegan et al, 2007 |
| BCD1 (amino acids 1-360) | pGEX-6P-2 | K.S.McKeegan | McKeegan et al, 2007 |

Table 2.11: Expression plasmids used to produce recombinant GST tagged proteins

N-C-fibrillarin (amino acids 81-321), 15.5K, TIP48 and TIP49 were also expressed from plasmids that introduced N terminal 6 x His fusion tags (see Table 2.12). Due to solubility issues recombinant deletion mutants of NOP56 (amino acids 1-458) and NOP58 (amino acids 1-435), which lacked the C terminal charged regions, had to be expressed. The C terminal charged region of NOP56 and NOP58, which was removed in the deletion mutants, has been shown not to be required for box C/D snoRNP

formation in yeast (Gautier et al, 1997; Lafontaine and Tollervey, 2000). The cDNA of NOP56 and NOP58 proteins was cloned into a pBAD/Thio-TOPO plasmid, which inserted an N terminal thioredoxin tag and C terminal His tag (See Table 2.11). The 6 x His tagged recombinant proteins were expressed and harvested as previously described for the GST tagged proteins; however, they were purified by immobilized metal affinity chromatography using His trap columns (GE healthcare). Immobilised 6 x His tagged proteins were eluted from the His trap columns using buffer A supplemented with 500 mM imidazole in an AKTA purification system (Amersham).

| Protein (all human) | Plasmid | Source | References |
|--------------------------------------|----------------|--------------|----------------------|
| N-C-fibrillarin (amino acids 82-321) | pET200 TOPO | K.S.McKeegan | N/A |
| 15.5K | pET15b | E.S.Maxwell | N?A |
| TIP48 | pET15b | E.S.Maxwell | Newman et al, 2000 |
| TIP49 | pET15b | E.S.Maxwell | Newman et al, 2000 |
| NOP56 (amino acids 1-458) | pBAD/Thio-TOPO | K.S.McKeegan | McKeegan et al, 2007 |
| NOP58 (amino acids 1-435) | pBAD/Thio-TOPO | K.S.McKeegan | McKeegan et al, 2007 |

Table 2.11: Expression plasmids used to produce recombinant 6 x His tagged proteins

2.7.2. Expression of *in vitro*-translated [³⁵S]methionine-labelled proteins

In vitro-translated ³⁵S-labelled proteins were expressed in the presence of [³⁵S] methionine in rabbit reticulocyte lysate using coupled *in vitro* transcription and translation kits (Promega), according to the manufactures' guidelines. To determine expression the *in vitro*-translated ³⁵S-labelled proteins were separated on a 12 % SDS-PAGE gel and visualised by autoradiography. *In vitro*-translated ³⁵S-labelled proteins were snap frozen in liquid nitrogen and stored at –80 °C. See Table 2.13 for details of expression plasmids.

| Protein (all human) | Plasmid | Source | References |
|--------------------------------------|-------------|--------------|--------------------------------|
| SMN | pET17b | U.Fischer | N/A |
| Gemin2 | pET17b | U.Fischer | Jones et al, 2001 |
| Gemin3 | pET17b | U.Fischer | N/A |
| Gemin4 | pET17b | U.Fischer | N/A |
| Gemin5 | pET17b | U.Fischer | N/A |
| Gemin6 | pET17b | U.Fischer | N/A |
| Gemin7 | pET17b | U.Fischer | N/A |
| Gemin8 | pET17b | U.Fischer | N/A |
| UNRIP | pET17b | U.Fischer | N/A |
| NUFIP | pcDNA | P.Cabart | Cabart et al, 2004 |
| TIP48 | pET15b | E.S.Maxwell | Newman et al, 2000 |
| TIP49 | pET15b | E.S.Maxwell | Newman et al, 2000 |
| TAF9 | pET17b | R.Tjian | Klemm et al, 1995 |
| NOP17 | pGEM-T easy | K.S.McKeegan | McKeegan et al, 2007 |
| PHAX | pET17b | K.S.McKeegan | McKeegan et al, 2009 (in prep) |
| fibrillarin | pET17b | C.M.Debieux | N/A |
| R-fibrillarin (amino acids 1-81) | pET17b | C.M.Debieux | N/A |
| R-N-fibrillarin (amino acids 1-141) | pET17b | C.M.Debieux | N/A |
| N-C-fibrillarin (amino acids 82-321) | pET17b | C.M.Debieux | N/A |
| C-fibrillarin (amino acids 142-321) | pET17b | C.M.Debieux | N/A |
| N-fibrillarin (amino acids 82-141) | pET100 | C.M.Debieux | N/A |

Table 2.13: Expression plasmids used to produce *in vitro*-translated ³⁵S-labelled proteins

2.7.3. *In vitro* protein-protein interaction assays

To perform a protein-protein *in vitro* interaction assay 20 µl of glutathione Sepharose beads (GE healthcare) were washed (3 x) in 1 ml cold buffer A (20 mM Tris-HCl [pH 8.0], 300 mM KCl, 0.1 % tween 20 (v/v), 10 % glycerol (v/v), 1 mM Tris-hydroxyphosphine) and pelleted by centrifugation, using a Eppendorf minifuge at 300 rpm for 1 minute. The washed beads were then incubated with the GST tagged protein of interest in a total volume of 200 µl with buffer A. After 2 hours at 4 °C, with constant agitation, the glutathione Sepharose beads were washed in cold buffer A (3 x) to remove unbound proteins (as described earlier). The protein of interest (recombinant His tagged protein or *in vitro*-translated ³⁵S-labelled protein) was added to the glutathione Sepharose beads, the sample made up to 200 µl with buffer A and incubated for 2 hours at 4 °C with constant agitation. Glutathione Sepharose beads (with immobilised proteins) were washed in buffer A (as described earlier) to remove unbound proteins, beads transferred to a new tube (to reduce background) and proteins removed using protein loading

buffer. Proteins were then separated on a 12 % SDS-PAGE gel and results visualised either by Coomassie blue staining, Western blot analysis or autoradiography.

2.7.4. *In vitro* protein-protein interaction assays using thioredoxin tagged proteins

To perform a protein-protein *in vitro* interaction assay using thioredoxin tagged proteins 10 µl of protein A Sepharose beads (GE healthcare) were washed (3 x) in buffer A and pelleted by centrifugation, using a Eppendorf minifuge at 300 rpm for 1 minute. The washed protein A sepharose beads were made up to 100 µl with buffer A and incubated with monoclonal anti-thioredoxin antibodies (Sigma, T0803) at 4 °C, with constant agitation. After 1 hour (or overnight) the protein A Sepharose beads were washed (3 x) with 1 ml of cold buffer A (as described earlier) and incubated with the thioredoxin tagged protein of interest at 4 °C, with constant agitation. After 2 hours the protein A Sepharose beads (containing immobilised antibodies and proteins) were washed with cold buffer A (as described above) and incubated with the *in vitro*-translated ³⁵S-labelled protein of interest in a total volume of 200 µl (with buffer A) at 4 °C, with constant agitation. After 2 hours the protein A Sepharose beads (containing antibodies and proteins) were washed (as described earlier), beads transferred to a new tube (to reduce background) and proteins removed from beads using protein loading buffer. Proteins were then separated on a 12 % SDS-PAGE gel and results visualised by autoradiography.

Chapter three

Development of approaches to analyse box C/D snoRNP biogenesis

3.1. Introduction

The box C/D snoRNPs are stable RNA protein complexes that function in the chemical modification of rRNA. There are over 100 different species of box C/D snoRNP that (Lestrade and Weber, 2006) in humans are responsible for over 105 specific rRNA 2'-*O*-methylation events, which are essential for the structure and function of the ribosome (Maden, 1990; Decatur and Fournier, 2002). A minority of box C/D snoRNPs, such as the U3 and U8 snoRNPs, do not function in the chemical modification of rRNA but rather function as chaperones assisting in pre-rRNA processing.

The box C/D snoRNPs consist of four evolutionary conserved common core proteins known as 15.5K, NOP56, NOP58 and fibrillarin (Tyc and Steitz, 1989; Lyman et al, 1999; Lafontaine & Tollervey, 1999; Lafontaine & Tollervey, 2000; Newman et al, 2000; Watkins et al, 2000). Fibrillarin is the methyltransferase component of the complex responsible for the 2'-*O*-methylation of target rRNA residues (Tollervey et al, 1993). The function of the other core proteins is not clear but they are most likely required for the structural integrity of the complex and their presence is vital for the biogenesis of box C/D snoRNPs (Lafontaine & Tollervey, 2000; Watkins et al, 2000, 2002). The box C/D snoRNA component of the complex contains conserved C (RUGAUGA, where R is purine) and D (CUGA) box motifs located at the 5' and 3' ends, respectively, which form the C/D box. Some box C/D snoRNAs also contain a second set of less conserved internal box motifs known as the C' and D' boxes. The snoRNA guides 2'-*O*-methylation by base pairing to complementary rRNA sequences using a 10-21 guide sequence located upstream of the D and / or D' box (Schimmang et al, 1989; Tyc and Steitz, 1989; Kiss-Laszlo et al, 1996; Lafontaine & Tollervey, 1999, 2000; Newman et al, 2000; Watkins et al, 2000).

The box C/D snoRNPs are found in the nucleolus where the synthesis, processing and modification of pre-rRNA transcripts occur as well as the bulk of ribosome assembly events. The nucleolus contains three sub-compartments, the FC, DFC and GC. As the pre-rRNA transcripts migrate through the nucleolus from the FC to the DFC and finally the GC they are processed and modified by factors that are specifically localised in these sub-compartments. The box C/D snoRNPs that function in the chemical

modification of rRNA as well as the box C/D snoRNPs that are involved in chaperoning early pre-rRNA processing, such as the U8 snoRNP (Peculis and Steitz, 1993), are found in the discrete DFC sub-compartments of the nucleolus (Figure 3.1). However, factors such as the U3 box C/D snoRNP, which function in both early and late pre-rRNA processing, are found throughout the nucleolus in both the DFC and GC (Figure 3.1; Granneman, et al, 2004). As the vast majority of box C/D snoRNPs are localised in the DFC of the nucleolus the core protein fibrillarin is found predominately in the DFC (Figure 3.1). In addition to the nucleolus the box C/D snoRNAs and fibrillarin are also found in the Cajal bodies and this, more than likely, represents pre-snoRNPs during biogenesis (Boulon et al, 2004; Lemm et al, 2006). The fibrillarin visualised in Cajal bodies may also represent fibrillarin that has been associated with the Cajal body localised scaRNPs (Jady and Kiss, 2001).

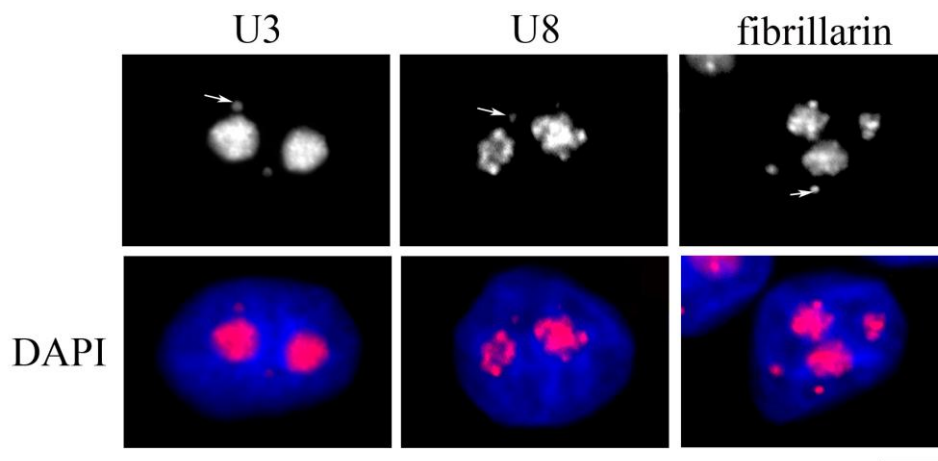


Figure 3.1: Nuclear localisation of the box C/D snoRNAs and fibrillarin

HeLa SS6 cells were grown on coverslips and fixed using paraformaldehyde. To detect box C/D snoRNA FISH was performed using fluorescent probes specific for the U3 and U8 snoRNAs. To determine fibrillarin localisation HeLa SS6 cells were transfected with a plasmid that expressed fibrillarin with an N terminal 2 x FLAG tag 6 x His tag and immunofluorescence performed using anti-FLAG antibodies. The top row shows the U3, U8 snoRNAs or fibrillarin while the bottom row is a merged image of U3, U8 snoRNAs or fibrillarin (red) with the DAPI image (blue). The white arrows indicate a Cajal body. The bar shown to the bottom right of the panel represents 5 μ m.

Box C/D snoRNP biogenesis occurs in a large complex known as the pre-snoRNP that consists of snoRNA, core proteins and assembly, RNA processing and transport factors (Watkins et al, 2004, 2007; McKeegan et al, 2007). The box C/D snoRNAs are transcribed from either independent genes or from the introns of housekeeping genes by

RNA polymerase II as extended pre-snoRNAs. The pre-snoRNAs are processed by numerous RNA processing factors to produce mature length snoRNA (Petfalski et al, 1998; Villa et al, 1998; Chanfreau et al, 1998; Allmang et al, 1999). The pre-snoRNA of independently transcribed box C/D snoRNAs, such as U3, also contains a 5' m⁷G cap that is hypermethylated by TGS1 to form an m₃G cap (Speckmann et al, 2000; Verheggen et al, 2002). The assembly factors TIP48, TIP49, BCD1, TAF9, NOP17 and NUFIP have been proposed to be involved in the early recruitment of the common core proteins to the pre-snoRNP (Peng et al, 2003; Watkins et al, 2004, 2007; Boulon et al, 2008; Gonzales et al, 2005; McKeegan et al, 2007).

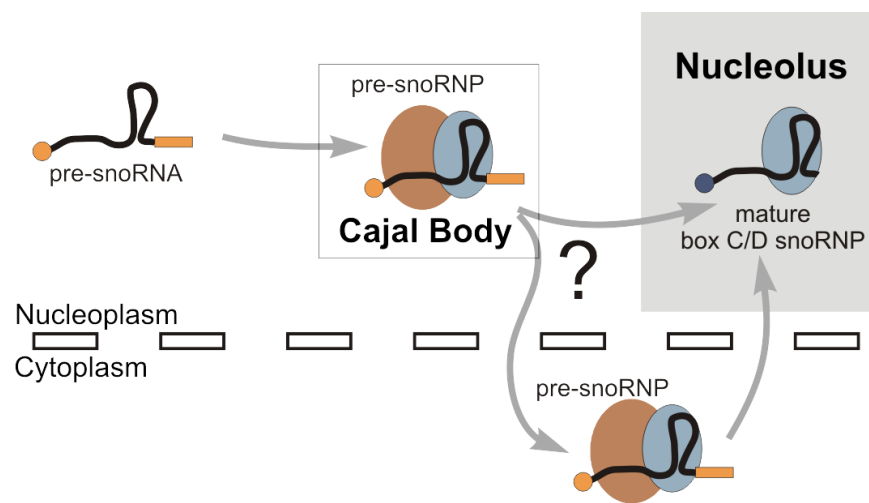


Figure 3.2: Box C/D snoRNP biogenesis

A black line represents the box C/D snoRNA with the 5' m⁷G cap shown as an orange circle and the 3' extended sequence as an orange rectangle. The core proteins are represented by a blue oval. The assembly, RNA processing and transport factors are represented by a brown oval. The 5' m₃G cap of mature snoRNA is shown as by a dark blue circle. The Cajal body is represented as a grey outlined box while the nucleolus is shown as a solid grey box. Grey arrows show the progression of biogenesis. The nuclear membrane is shown by black outlined rectangles. Image provided by N.J.Watkins, Newcastle University, UK.

The cellular localisation(s) where box C/D snoRNP biogenesis occurs is not clear with evidence suggesting the nucleoplasm (Terns et al, 1995; Watkins et al, 2004; Boulon et al, 2004), Cajal bodies (Verheggen et al, 2001; Boulon et al, 2004; Lemm et al, 2006) and cytoplasm (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007). Numerous transport factors have been found associated with pre-snoRNPs such as PHAX, CRM1, Ran, CBC, Nopp140 and Snurportin1 (Yang et al, 2000; Boulon et al, 2004; Watkins et

al, 2004, 2007). These factors, more than likely, transfer the box C/D pre-snoRNPs to sub-cellular domains where snoRNP biogenesis occurs.

To investigate box C/D snoRNP biogenesis it was necessary to establish a variety of experimental assays, procedures and approaches. To determine the cellular location of box C/D snoRNP biogenesis FISH and immunofluorescence were used to identify the localisation of snoRNA and the box C/D snoRNP core proteins and assembly factors. These assays were also used in conjunction with siRNA mediated depletion to determine the effect of the loss of a factor of interest on localisation and biogenesis of box C/D snoRNPs. An issue that limited the use of these approaches was that it was not possible to discriminate between mature snoRNPs, which had yet to be turned over, and snoRNPs that had been synthesised after the siRNA mediated depletion of a factor of interest. To address this issue a number of approaches were developed that specifically detected newly synthesised box C/D snoRNPs. These included the use of reporter plasmids and the development of two box C/D snoRNA inducible cell lines. In addition three inducible protein systems were developed, which expressed tagged fibrillarin and SMN proteins. Using these systems, in conjunction with siRNA mediated depletion, the analysis of box C/D snoRNPs which had been synthesised in the absence of a factor of interest was possible.

3.2. Results

3.2.1. The role of the common core proteins in box C/D snoRNP localisation

Previous studies have shown that loss of the box C/D snoRNP common core proteins NOP56, NOP58 and fibrillarin results in a reduction of box C/D snoRNA levels, which represents a decrease in snoRNP biogenesis (Lafontaine and Tollervey, 2000; Watkins et al, 2004). To further characterise the effect of the loss of the common core proteins on box C/D snoRNP biogenesis siRNAs targeting NOP56, NOP58 and fibrillarin mRNA were used to deplete the levels of the respective proteins and the localisation of box C/D snoRNA analysed by FISH. However, first it was necessary to analyse the efficiency of the siRNAs in reducing target protein levels.

The siRNA duplexes used to deplete NOP56, NOP58 and fibrillarin have been successfully used by Watkins et al 2004; however, to confirm their efficiency HeLa SS6 cells were transfected with siRNAs targeting fibrillarin mRNA and the depletion of the respective protein analysed. As a control a parallel experiment was performed in which HeLa SS6 cells were transfected with a control siRNA. The control siRNA used targeted firefly luciferase mRNA, which is not present in human cells, and therefore provided a control of the transfection procedure (Elbashir et al, 2002). Sixty hours after transfection cells were harvested, proteins separated on a 12 % SDS-PAGE gel and Western blot assays performed using antibodies specific for fibrillarin and nucleolin. The nucleolin protein levels were used to control sample loading and to determine the specificity of the fibrillarin depletion.

In cells transfected with siRNAs targeting fibrillarin mRNA there was a dramatic reduction of fibrillarin protein levels (Figure 3.3), which was consistent with the results of Watkins et al 2004. The effect of the siRNAs was specific as the levels of nucleolin were unaffected. The specificity and the efficiency of the siRNAs targeting NOP56 and NOP58 mRNA have all been tested in Watkins et al, 2004 and, as with the fibrillarin siRNAs, result in the specific depletion of target protein levels.

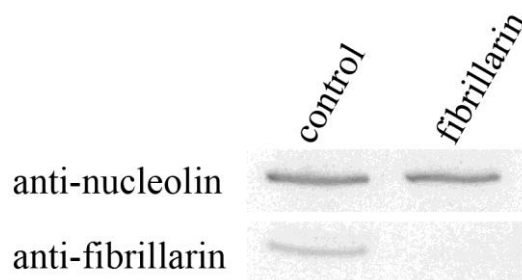


Figure 3.3: Depletion of fibrillarin

HeLa SS6 cells were transfected with siRNAs targeting either firefly luciferase (control) or fibrillarin mRNA. Cells were harvested sixty hours after transfection, proteins separated on a 12 % SDS-PAGE gel and Western blot assays performed using antibodies that recognise fibrillarin and nucleolin. Nucleolin provided a control of sample loading as well as the specificity of the siRNA depletion. The proteins targeted for depletion are indicated above the panel. The antibodies used are indicated to the left of the panel.

To determine the effect of the depletion of the common core proteins on box C/D snoRNA localisation siRNAs targeting either NOP56, NOP58, fibrillarin or firefly luciferase (control) were transfected into HeLa SS6 cells. Sixty hours after transfection the cells were fixed and FISH performed using fluorescent probes specific for the U3 box C/D snoRNA and the spliceosomal U2 snRNA. As the U2 snRNA is found in nucleoplasmic Cajal bodies and nuclear speckles it provided a marker of these nuclear bodies. The U2 snRNA also provided a control of the specificity of the siRNA depletion as none of the factors targeted for depletion are known to be involved in snRNP biogenesis and therefore the loss of these factors should only affect the U3 snoRNA.

In the control cells the U3 snoRNA was found predominately in the nucleolus but was also found in Cajal bodies (Figure 3.4), the typical localisation of U3 box C/D snoRNA. The U3 snoRNA was distributed throughout the nucleolus indicating localisation to both the DFC and GC (see Figure 3.1 for comparison of nucleolar localisation). The distribution and / or level of U2 snRNA did not change with any of the siRNA depletions, which indicates that all the effects seen were specific to the U3 snoRNAs.

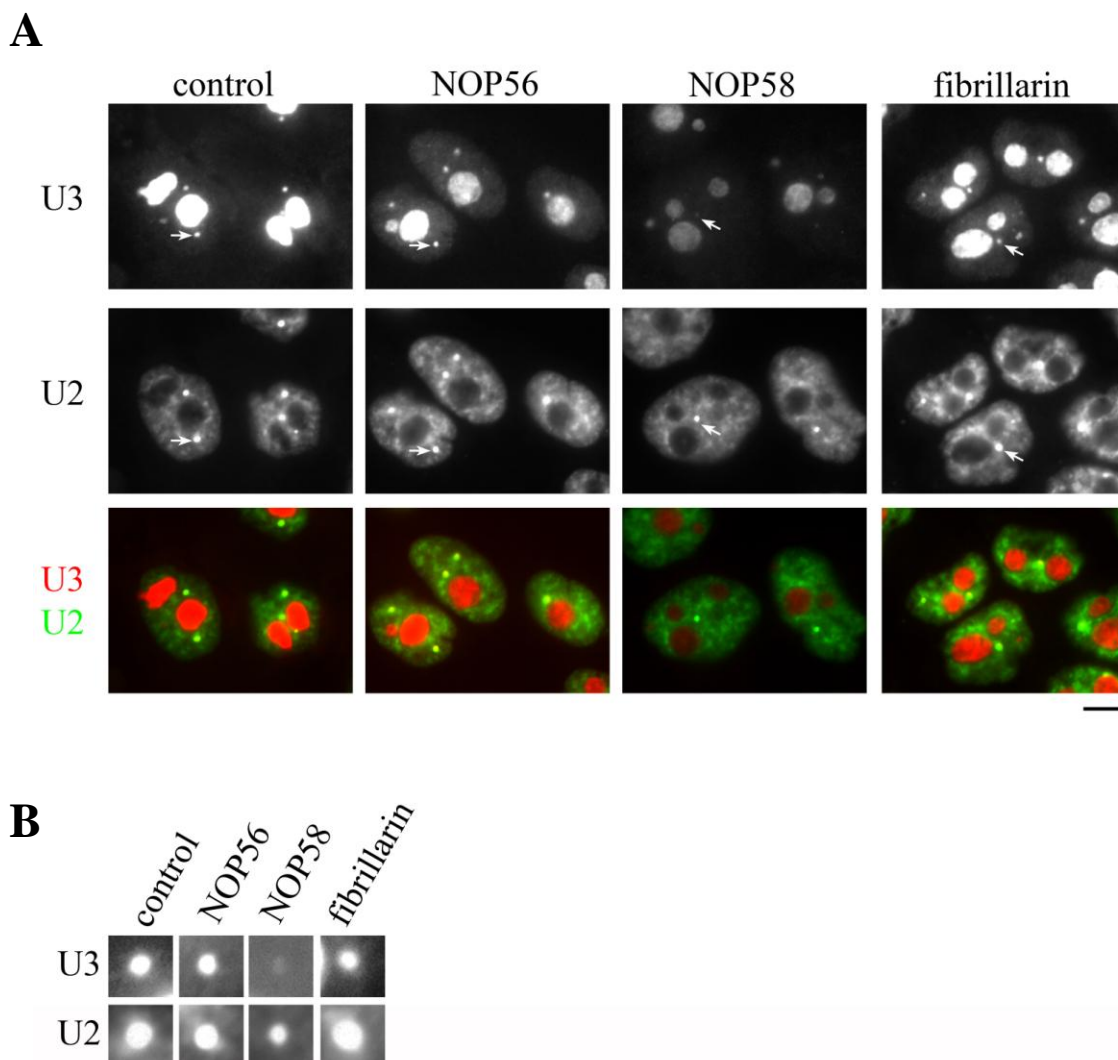


Figure 3.4: The localisation of box C/D snoRNA in cells depleted of the common core proteins

(A) HeLa SS6 cells were transfected with siRNAs targeting either firefly luciferase (control) NOP56, NOP58 or fibrillarin. After sixty hours cells were fixed and FISH performed using fluorescent probes specific for the U3 box C/D snoRNA and U2 snRNA. The same exposure time was used for each probe to allow direct comparison of RNA levels and distribution. The protein targeted for depletion is indicated at the top of each column of images. The labels along the left hand side of the panels indicate the FISH probes used. The bottom row shows a merged image of the U3 snoRNA (red) and U2 snRNA (green). The white arrows show a selected Cajal body. The bar to the bottom right of the panel represents 5 μ m.

(B) An enlargement of a Cajal body from each siRNA depleted cell from Figure 3.4 A. The protein targeted for depletion is indicated at the top of each column of images. The labels along left hand side of panels indicate FISH probes used.

In both NOP56 and fibrillarin depleted cells there was a mild reduction in the levels of the nucleolar localised U3 snoRNA compared to the control cells (Figure 3.4). However, the relative intensity of the U3 snoRNA signal from the Cajal bodies remained unchanged in NOP56 and fibrillarin depleted cells compared to the control cells, indicating a retention or accumulation of the U3 snoRNA. The depletion of NOP56 and fibrillarin also resulted in the accumulation of the U3 snoRNA in the nucleoplasm of the cell. The accumulation of the U3 snoRNA in the nucleoplasm and Cajal bodies, more than likely, corresponds to a halt in box C/D snoRNP biogenesis. A possible explanation is that pre-snoRNPs were formed, which lacked either NOP56 or fibrillarin, and were unable to localise to the nucleolus resulting in their accumulation in the nucleoplasm and Cajal bodies. The reduction in nucleolar U3 snoRNA levels most likely reflects a decrease in box C/D snoRNP biogenesis.

In cells depleted of NOP58 there was a dramatic reduction in U3 snoRNA levels in the nucleolus and Cajal bodies (Figure 3.4). The drop in the U3 snoRNA levels as a result of the depletion of NOP58 was, most likely, due to NOP58 being essential for the stability of pre-snoRNAs as has been previously proposed (Lafontaine and Tollervey, 2000; Watkins et al, 2004).

3.2.2. Loss of the unrelated protein, hPRP31, does not affect box C/D snoRNA localisation

To ensure that the change in the U3 snoRNA localisation and accumulation seen when NOP56, NOP58 and fibrillarin were depleted (Figure 3.4) was not due to an indirect effect, caused by the loss of an essential protein, an unrelated essential factor was depleted and the localisation of the U3 snoRNA analysed by FISH.

The protein targeted for depletion was hPRP31 (61K), a snRNP protein essential for the interaction between the U4/U6 di-snRNP and the U5 snRNP to form the U4/U6.U5 tri-snRNP. The U4/U6.U5 tri-snRNPs are distributed between nuclear speckles and Cajal bodies and previous studies have shown that siRNA mediated depletion of hPRP31 results in the loss of the U4 and U6 snRNPs from nuclear speckles and an accumulation in the Cajal bodies (Schaffert et al, 2004). From this observation it was proposed that the Cajal bodies were the site where and U4/U6.U5 tri-snRNP is formed (Schaffert et al, 2004).

HeLa SS6 cells, grown on coverslips, were transfected with siRNAs targeting either firefly luciferase (control), NOP58 or hPRP31. Sixty hours later cells were fixed and analysed by FISH using fluorescent probes specific for the U3 box C/D snoRNA and the spliceosomal U4 snRNA (Figure 3.5). In the control cells the U3 snoRNA was localised predominately in the nucleolus but also in the Cajal bodies while the U4 snRNA was distributed between the nuclear speckles and Cajal bodies. Cells transfected with siRNAs targeting NOP58 showed a dramatic reduction in the U3 snoRNA levels (as in Figure 3.4) but the U4 snRNA level and distribution were not affected. When the U4/U6 snRNP specific hPRP31 protein was depleted the U4 snRNA accumulated in Cajal bodies, as was previously described by Schaffert et al 2004. Importantly when hPRP31 was depleted there was no change in the localisation or levels of the U3 snoRNA. This indicated that the effects seen with the U3 snoRNA localisation and distribution, when NOP56, NOP58 and fibrillarin were depleted, were specific to the functions of these proteins.

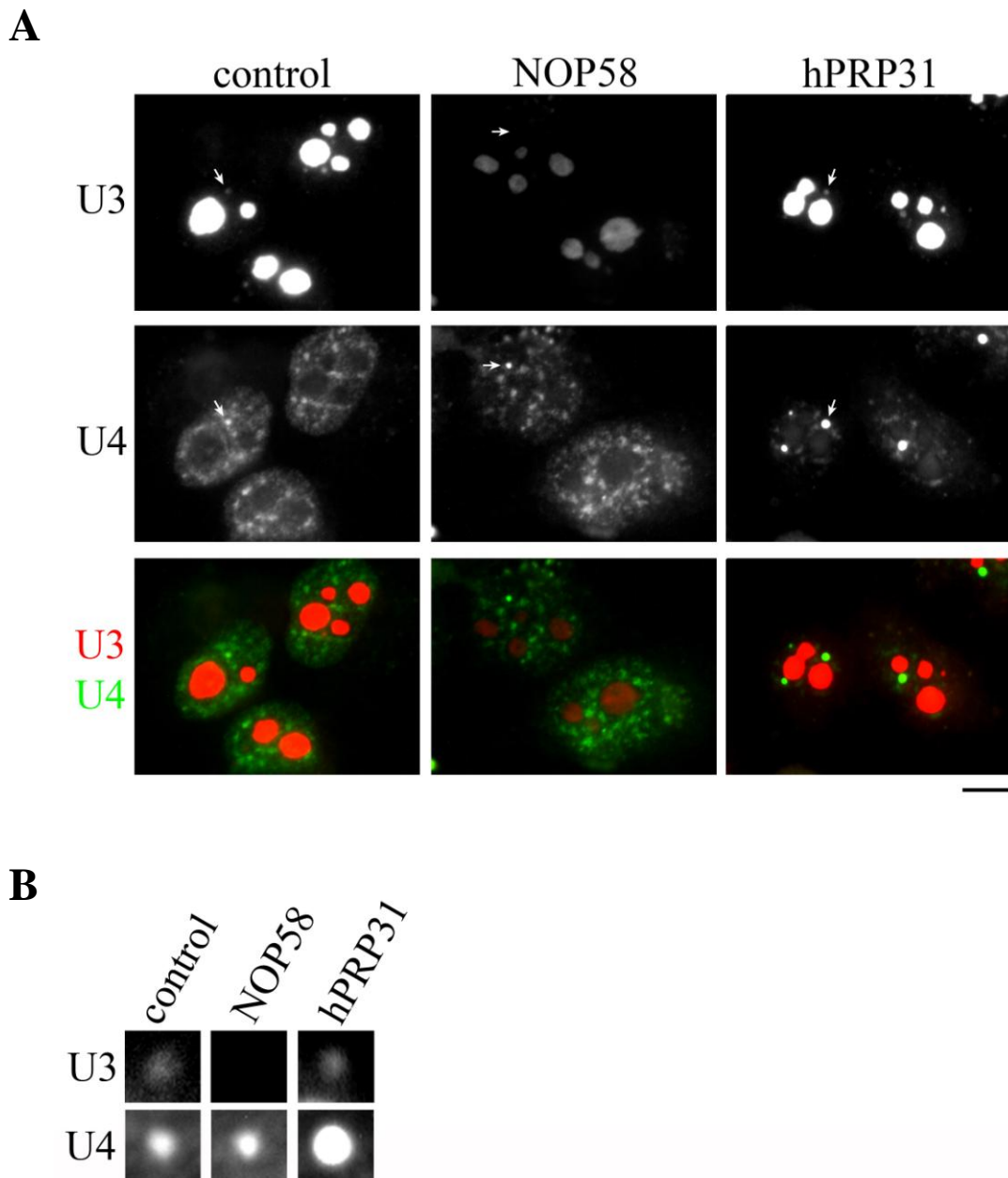


Figure 3.5: The localisation of box C/D snoRNA in cells depleted of hPRP31

(A) HeLa SS6 cells were transfected with siRNAs targeting firefly luciferase (control), NOP58 or hPRP31. After sixty hours cells were fixed and FISH performed using fluorescent probes specific for the U3 box C/D snoRNA and the spliceosomal U4 snRNA. The same exposure time was used for each probe to allow direct comparison of RNA levels and distribution. The protein targeted for depletion is indicated at the top of each column of images. The labels along the left hand side of the panels indicate the FISH probes used. The bottom row shows a merged image of the U3 snoRNA (red) and the U4 snRNA (green). The white arrows indicate a selected Cajal body. The bar shown to the bottom right of the panel represents 5 μ m.

(B) An enlargement of a Cajal body from each siRNA depleted cell from Figure 3.4 A. The protein targeted for depletion is indicated at the top of each column of images. The labels along left hand side of panels indicate FISH probes used.

3.2.3. Localisation of the box C/D snoRNP assembly factors

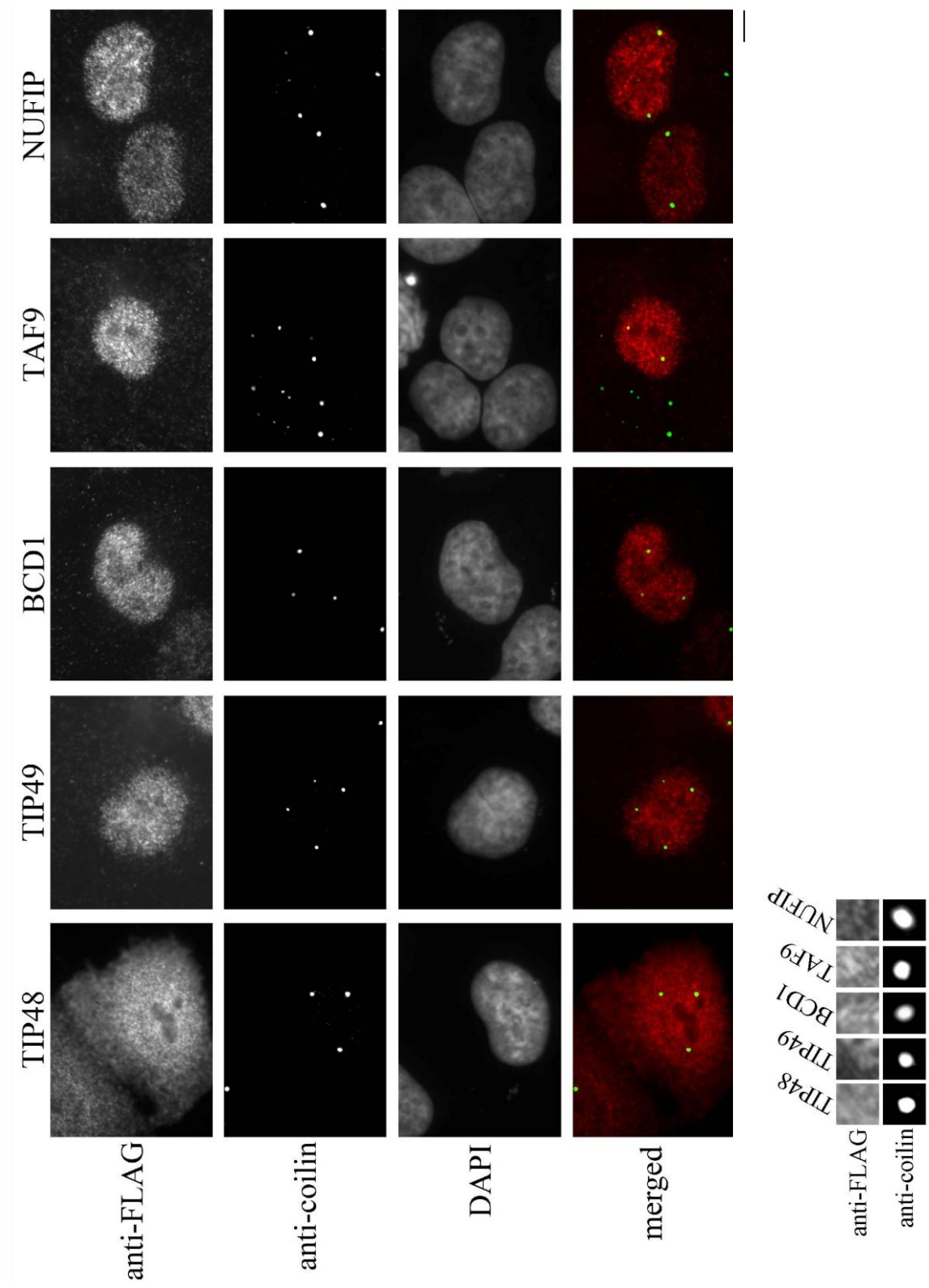
Numerous factors have been implicated in the assembly of the box C/D snoRNPs; however, the localisation of many of these factors has not been analysed. The identification of the localisation of box C/D snoRNP assembly factors in comparison with the distribution of the box C/D snoRNA and common core proteins may shed light on the site(s) of box C/D snoRNP biogenesis.

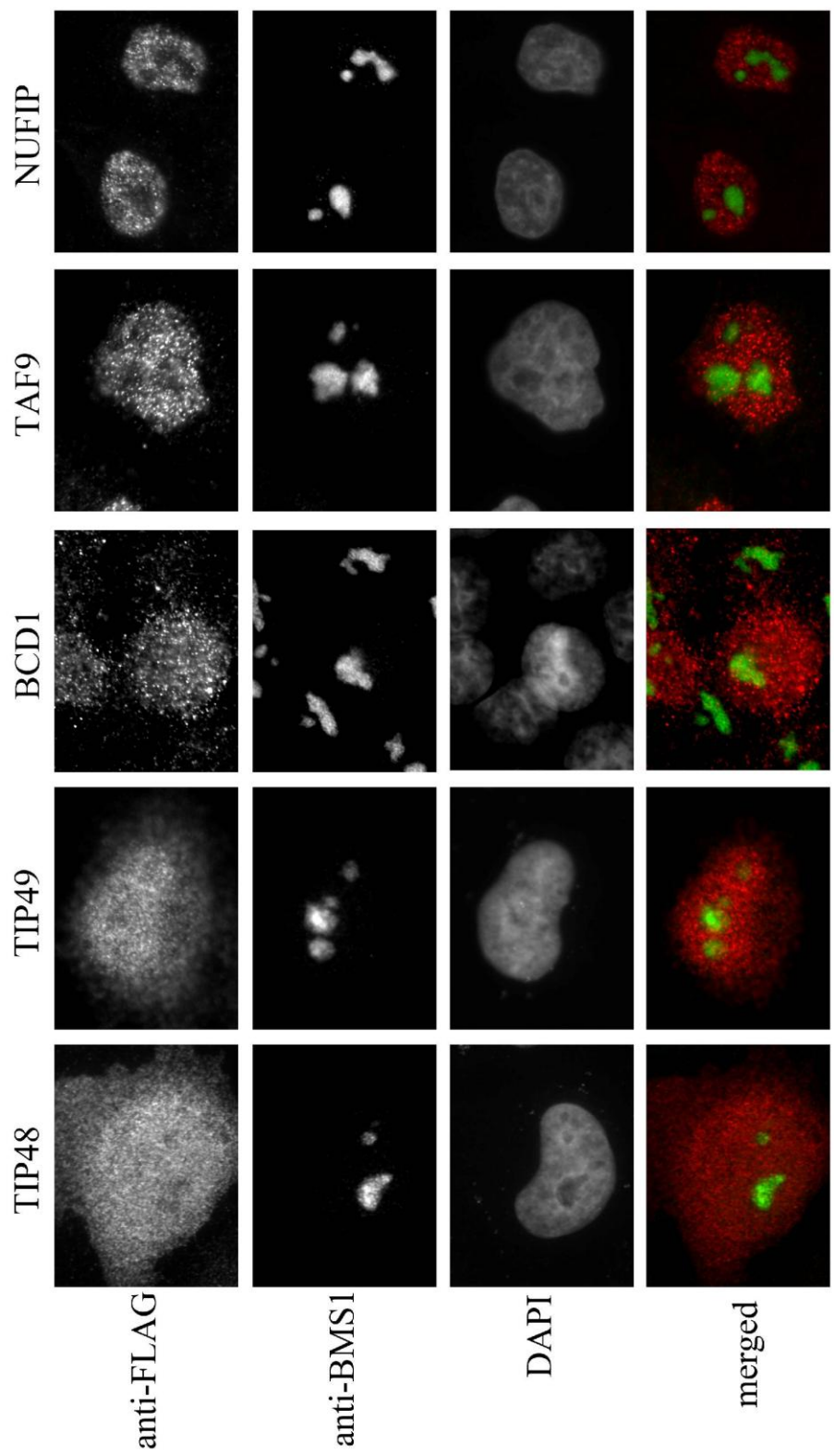
As antibodies were not available for a number of the box C/D snoRNP assembly factors plasmids were constructed that expressed the assembly factors fused to an N terminal 2 x FLAG tag (plasmids provided by K.S.McKeegan, Newcastle University, UK). HeLa SS6 cells, grown on coverslips, were transfected with the expression plasmids and twenty-four hours later cells were fixed. The localisation of the assembly factors was analysed by immunofluorescence using anti-FLAG antibodies. Cells were also incubated with either anti-coilin or anti-BMS1 antibodies to detect the Cajal bodies or nucleolus, respectively. DAPI staining of DNA was used to determine the localisation of the nucleoplasm.

The box C/D snoRNP assembly factors TIP48, TIP49, BCD1, TAF9 and NUFIP, expressed from the plasmids, were found distributed throughout the nucleoplasm (Figure 3.6 A and B). In addition TIP48 was found in the cytoplasm. None of these assembly factors were found enriched in the Cajal bodies, which were indicated by coilin (Figure 3.6 A). The localisation of assembly factors in the Cajal bodies cannot, however, be excluded as there is no definitive absence of the factors from these nuclear bodies. Antibodies raised against BMS1 were used to identify the nucleolus but the localisation of assembly factors to this region was not clear-cut (Figure 3.6 B). Cells transfected with the NUFIP, BCD1 and TAF9 expression plasmids showed a weak nucleolar staining using the anti-FLAG antibody but this was much lower than the nucleoplasmic staining and not repeatedly seen. The low levels of assembly factors sometimes observed in the nucleolus may be due to some cells being transfected with large quantities of the expression plasmids. This could result in over expression and accumulation of tagged protein to areas of the cell where it does not normally localise. Overall the data indicates that the assembly factors localise to the nucleoplasm.

Figure 3.6: The localisation of the box C/D snoRNP assembly factors

HeLa SS6 cells were transfected with plasmids expressing box C/D snoRNP assembly factors fused to N terminal 2 x FLAG tags. Twenty-four hours after transfection HeLa SS6 cells were fixed and immunofluorescence performed using anti-FLAG antibodies to determine the location of the assembly factors. In addition immunofluorescence was performed with **(A)** anti-coilin antibodies or **(B)** anti-BMS1 antibodies to indicate the locations of the Cajal bodies and nucleolus, respectively. DAPI staining of DNA was used to identify the nucleoplasm. The assembly factor transfected is indicated at the top of each column of images. Labels along the left hand side of the panels indicate antibodies used and DAPI staining. The bottom row shows merged image of FLAG tagged assembly factors (red) with coilin or BMS1 (green), as indicated. The bar shown to the bottom right of the panel represents 5 μ m. The BCD1, TAF9 and NUFIP images with BMS1 were kindly provided by K.S.McKeegan, Newcastle University, UK. **(A)** Image below main panel is an enlargement of a Cajal body from each cell of Figure 3.6 A. The assembly factor transfected is indicated at the top of each column of images. Labels along the left hand side of the panels indicate antibodies used.





3.2.4. Development of procedures to analyse *de novo* box C/D snoRNP biogenesis in HeLa SS6 cells

One issue with using siRNAs to deplete a protein of interest followed by the analysis of the localisation of endogenous box C/D snoRNA was that it was not possible to discriminate between snoRNAs synthesised prior to protein depletion and newly formed snoRNAs. As box C/D snoRNPs are predicted to be as stable as the snRNPs (Personal communication from N.J.Watkins, Newcastle University, UK), which have half-lives of between 60 and 88 hours (Fury and Zieve, 1996), it is likely that a high background of mature box C/D snoRNPs, synthesised prior to siRNA mediated depletion, are present. Due to these issues it was necessary to develop alternative approaches to analyse box C/D snoRNP biogenesis.

Previous studies investigating the synthesis of the snRNPs overcame the issues of the stability and long half-lives of the snRNPs by developing assays to monitor *de novo* biogenesis. One such study by Shpargel and Matera 2005 used a procedure whereby siRNAs were first used to deplete potential snRNP biogenesis factors, in this case SMN, then forty-eight hours later the cells were transfected with a GFP-SmB reporter plasmid. Analysis of the tagged protein derived from the reporter plasmid ensured that only *de novo* synthesised snRNPs, produced in the absence of SMN were detected. Using this approach, Shpargel and Matera 2005 showed that when SMN protein levels were depleted the GFP-SmB protein was unable to localise to Cajal bodies. Further experiments using this two-step transfection procedure also revealed that in the absence of SMN there was reduced binding of GFP-SmB to the snRNAs (Shpargel and Matera, 2005). Both these experiments supported the role for SMN in snRNP biogenesis. Interestingly, Shpargel and Matera 2005 were unable to detect any change in the localisation of the endogenous Sm proteins in SMN depleted cells. Therefore, the two step-transfection procedure, using a reporter plasmid to monitor *de novo* biogenesis, provided more reliable and accurate data. With this in mind, several procedures were developed to analyse *de novo* box C/D snoRNP biogenesis.

One of the procedures developed to analyse *de novo* box C/D snoRNP biogenesis utilised a U3 reporter plasmid (StreptoTag U3, Granneman et al, 2004). This plasmid expressed human U3 snoRNA with a Strepto tag in a non-essential 3' stem loop region of RNA (Figure 3.7). The StreptoTag U3 was detectable by FISH and primer extension assays using probes specific for the StreptoTag sequence (Granneman et al, 2004). The StreptoTag U3 reporter plasmid was incorporated into a two-step transfection procedure similar to one used by Shpargel and Matera 2005. In this procedure HeLa SS6 cells were transfected with siRNAs to deplete a target protein followed by a secondary transfection of the StreptoTag U3 reporter plasmid. This procedure allowed the *de novo* biogenesis of U3 snoRNPs to be specifically monitored in the absence of the protein of interest.

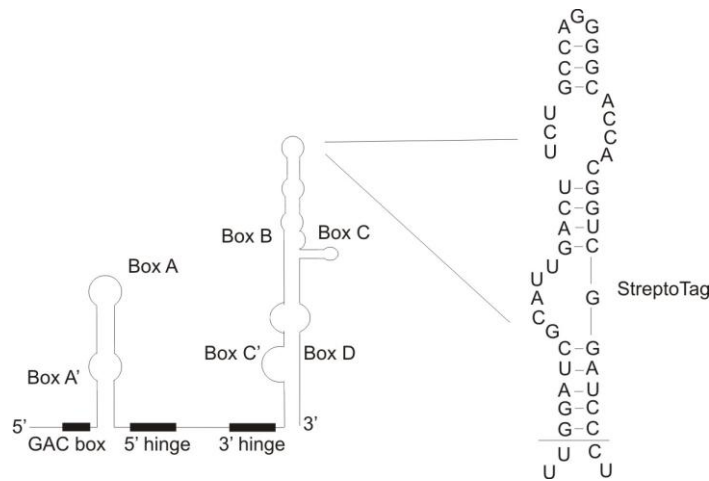


Figure 3.7: StreptoTag U3 snoRNA

Schematic diagram showing location and predicted secondary structure of the U3 box C/D snoRNA StreptoTag. The snoRNA is represented as a black line in the left panel. In the right panel the snoRNA is shown as black lettering. The locations of GAC box, 5' hinge and 3' hinge regions shown by black box. Conserved box A, A', C', D, B and C are indicated (Granneman et al, 2004).

To establish the two-step transfection procedure the effect of the loss of NOP56 and NOP58 on the localisation of newly synthesised StreptoTag U3 snoRNA was analysed. HeLa SS6 cells were transfected with siRNAs targeting either firefly luciferase (control), NOP56 or NOP58, and after twenty-four hours transfected with the StreptoTag U3 reporter plasmid. Twelve hours later cells were fixed and FISH performed using a fluorescent probe specific for the StreptoTag sequence to detect the

newly synthesised U3 snoRNA. As a comparison FISH was also performed using a fluorescent probe targeting endogenous U3 snoRNA.

In the control cells both the endogenous U3 and the StreptoTag U3 snoRNA were found predominately in the nucleolus but also in Cajal bodies, the typical localisations of box C/D snoRNAs (Figure 3.8). Endogenous U3 and StreptoTag U3 snoRNAs were distributed throughout the nucleolus indicating localisation to both the DFC and GC, which was consistent with previous studies (Granneman et al, 2004; see Figure 3.1 for comparison of nucleolar localisation). In cells depleted of NOP56 there was an accumulation of both the endogenous U3 and the StreptoTag U3 snoRNAs in the nucleoplasm. This was, most likely, due to a block in biogenesis with pre-snoRNP complexes being formed, which lacked NOP56, resulting in nucleoplasmic accumulation. The StreptoTag U3 snoRNA was also present in Cajal bodies in NOP56 depleted cells; however, it is difficult to determine whether there was an accumulation in these bodies without a transfection control. Interestingly, some StreptoTag U3 snoRNA also accumulated in the nucleolus in NOP56 depleted cells. This indicated that either there was a low level production of mature U3 particles in the reduced NOP56 environment or that snoRNP complexes were formed that lacked NOP56 but maintained the ability to localise, albeit inefficiently, to the nucleolus. In NOP56 depleted cells there was also a reduction in the levels of nucleolar endogenous U3 snoRNA.

In NOP58 depleted cells there was a dramatic reduction of endogenous U3 levels with no visible accumulation in the nucleoplasm or Cajal bodies (Figure 3.8), as seen before (Figure 3.4). The StreptoTag U3 expressed in the NOP58 depleted cells did accumulate in low levels in the nucleolus but not in either the nucleoplasm or Cajal bodies (Figure 3.8); however, without a transfection control it is difficult to interpret this data.

Further two-step transfection experiments were performed using the StreptoTag U3 reporter plasmid and siRNAs targeting fibrillarin and a variety of other assembly factors. It was, however, not possible to perform a successful secondary transfection of the StreptoTag U3 reporter plasmid. One possible reason for this is that the loss of an essential protein may affect the transfection rate of the cells. To overcome the issue of poor secondary transfection different methods were tested including electroporation; however, these attempts were also unsuccessful.

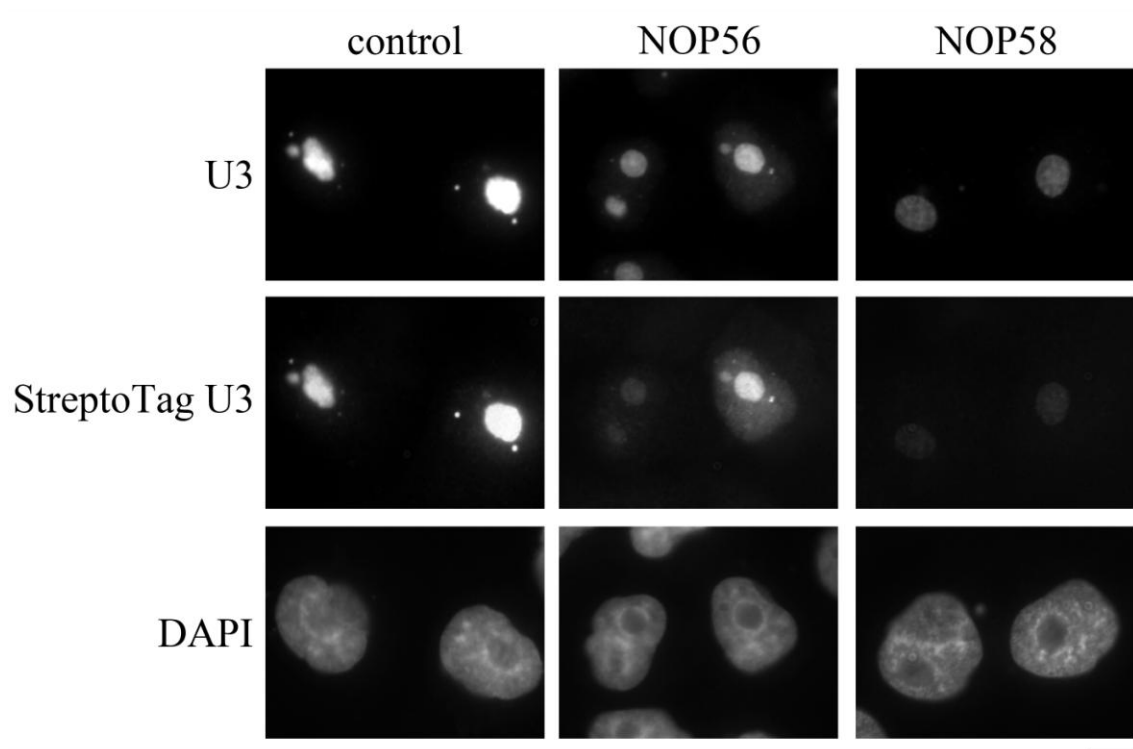


Figure 3.8: The localisation of newly synthesised StreptoTag U3 snoRNA in cells depleted of NOP56 and NOP58

HeLa SS6 cells were transfected with siRNAs targeting either firefly luciferase (control), NOP56 or NOP58. Twenty-four hours later cells were transfected with a reporter plasmid expressing U3 snoRNA tagged with strepavidin (StreptoTag U3). After a total of sixty hours, cells were fixed and analysed by FISH using fluorescent probes specific for the U3 snoRNA and the StreptoTag U3 snoRNA. DAPI staining of DNA was used to indicate the nucleoplasm. The same exposure time was used for each probe to allow direct comparison of RNA levels and distribution. The protein targeted for depletion is indicated at the top of each column of images. The labels along the left hand side of the panels indicate the FISH probes and DAPI staining. The bar to the bottom right of the panel represents 5 μ m.

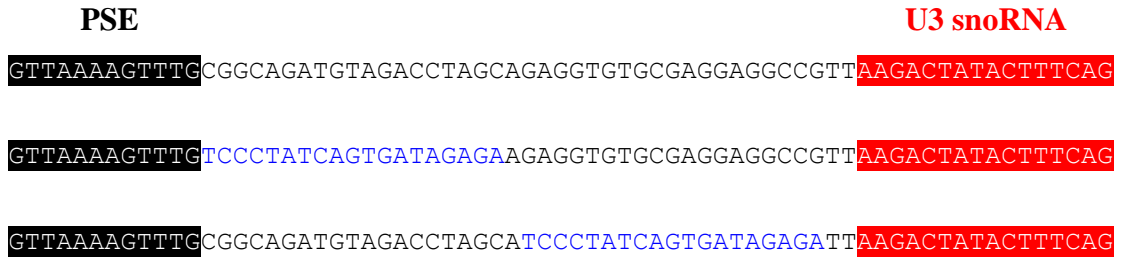
3.2.5. Development of inducible box C/D snoRNA cell lines

There was a clear problem with the two-step transfection procedure used to analyse *de novo* box C/D snoRNP biogenesis. In order to remove the need for a second transfection Tet-inducible cell lines were developed that expressed either the U3 or the U14 box C/D snoRNAs.

A Tet-inducible U3 HEK293 cell line was produced utilising the StreptoTag U3 reporter plasmid used previously in the two-step transfection procedure (Figure 3.7). The use of StreptoTag U3 snoRNA meant that induced U3 snoRNA could be detected above endogenous U3 snoRNA by FISH and primer extension assays using probes specific for the StreptoTag sequence (Granneman et al, 2004). As the U3 snoRNA is transcribed from an independent gene a Tet₀₂ sequence (5'-TCCCTATCAGTGATAGAG A-3') was inserted into the U3 promoter region. The development of the inducible U3 promoter was based upon a Tet-responsive U6 promoter that was constructed to express shRNAs (short hairpin RNAs; Lin et al, 2004). The Tet-responsive U6 promoter provided a good basis for the development of the Tet-responsive U3 promoter as the U3 snoRNA and U6 snRNA promoter regions contain similar transcription elements. A single Tet₀₂ sequence was inserted in a non-essential region of the human U3 promoter between the 5' PSE (Ach and Weiner, 1991; Savino et al, 1992) and the transcription start site (Figure 3.9). Two Tet-responsive StreptoTag U3 promoters were created that differed slightly in the location of the Tet₀₂ sequence.

Before creating the inducible HEK293 cell line it was first necessary to determine whether the insertion of the Tet₀₂ sequence had damaged the function of either of the new U3 snoRNA promoters. HeLa SS6 cells were transfected with the StreptoTag U3 reporter plasmids, containing the inserted Tet₀₂ sequences, and expression determined by primer extension assays. As HeLa SS6 cells do not contain the Tet-repressor gene the Tet-responsive promoters would not be repressed. These assays showed that only one of the StreptoTag U3 reporter plasmids, containing inserted Tet₀₂ sequences, retained the ability to express StreptoTag U3 (data not shown).

A



B



Figure 3.9: Schematic diagram of the Tet-responsive StreptoTag U3 snoRNA

(A) To create a Tet-responsive StreptoTag U3 promoter a single TetO₂ sequence (blue) was inserted between the PSE (proximal sequence element; black) and StreptoTag U3 sequence (red). The top panel shows a section of the U3 box C/D snoRNA gene with the PSE (proximal sequence element; white lettering on black background), U3 snoRNA sequence (white lettering on red background) and intervening sequences (black lettering). Middle and bottom panels show the location of the inserted TetO₂ operator sequences (Blue text).

(B) Schematic diagram of the location of StreptoTagU3 gene in an integrated pcDNA5/FRT/TO expression cassette. The FRT recombination sites mark the insertion site of pcDNA5/FRT/TO expression plasmid in the Flp-In T-Rex HEK293 cells. As the StreptoTag U3 gene contains its own promoter with inserted TetO₂ sequence it was inserted upstream of the CMV/2xTetO₂ promoter region between the ampicillin gene (Amp) gene and pUC origin. P_{SV40} represents Simian virus 40 promoter. BGH pA represents Bovine growth hormone polyadenylation signal. For further details of plasmid see Chapter 2.2.1.

To create a Tet-inducible StreptoTag U3 snoRNA cell line the StreptoTag U3 gene, with inserted TetO₂ sequence, was cloned into a pcDNA5/FRT/TO expression plasmid containing a FRT recombination site. The pcDNA5/FRT/TO expression plasmid was then inserted into the FRT recombination site of Flp-In T-Rex HEK293 cells through homologous recombination. The Flp-In T-Rex HEK293 cells contain a Tet-repressor gene that expresses TetR proteins that directly bind TetO₂ sequences and prevent their expression. The addition of Tet inhibits the function of the TetR proteins and therefore induces expression from promoters containing TetO₂ sequences.

To test the expression of StreptoTag U3 snoRNA from the Tet-inducible StreptoTag U3 HEK293 cell line varying quantities of Tet was added to the cells and primer extension assays performed using primers specific for the StreptoTag sequence. Unfortunately, only very low levels of StreptoTag U3 were detected, which were not sufficient for this cell line to be used in further experiments (data not shown).

In addition to the Tet-inducible StreptoTag U3 HEK293 cell line a Tet-inducible mouse U14 (mU14) HEK293 cell line was constructed. The induced mU14 snoRNA can be specifically detected above endogenous U14 snoRNA by primer extension assays due to sequence differences between mouse and human U14. As U14 is an intronic RNA, encoded within the introns of the hsc70 gene, to create the Tet-inducible mU14 it was not necessary to design a new promoter region, as the Tet-responsive CMV promoter contained within the pcDNA5/FRT/TO expression plasmid was suitable. A section of the mouse hsc70 gene containing exons 5 and 6 with intron 5 (containing U14) was cloned from a mouse U14 expression plasmid (Watkins et al, 1996) and inserted into a pcDNA5/FRT/TO expression plasmid (Figure 3.10 A). The cell line was then created as described earlier.

To determine the level of mU14 snoRNA expression from the Tet-inducible mU14 HEK293 cell line various concentrations of Tet were added to the cells and, after forty-eight or seventy-two hours, cells harvested. RNA extraction was performed on the samples followed by primer extension assays using primers specific for mU14 snoRNA. As a control of sample loading a primer extension assay was also performed with each sample using a primer specific for human U3 snoRNA. The samples were then separated by electrophoresis on an 8 % acrylamide gel containing 7 M urea and results viewed by autoradiography (Figure 3.10 B).

The U3 primer extensions performed show that equal quantities of RNA were loaded from each sample (Figure 10 B). In the absence of Tet no mU14 snoRNA was expressed (Figure 10 B; lane 1); however, when the mU14 HEK293 cells were grown in the presence of 2 µg/ml of Tet for forty-eight or seventy-two hours (Figure 3.10; lanes 2 and 3) a weak signal was present which indicates that there was a low level of mU14 snoRNA expression. When cells were grown in a lower concentration of Tet (1 µg/ml; Figure 3.10 B; lane 4 and 5) it took at least seventy-two hours for a detectable amount

of mU14 snoRNA to be expressed. Unfortunately, the expression levels of mU14 were very low and therefore the Tet-inducible mU14 HEK293 cells were not suitable for use in assays to analyse box C/D snoRNP biogenesis.

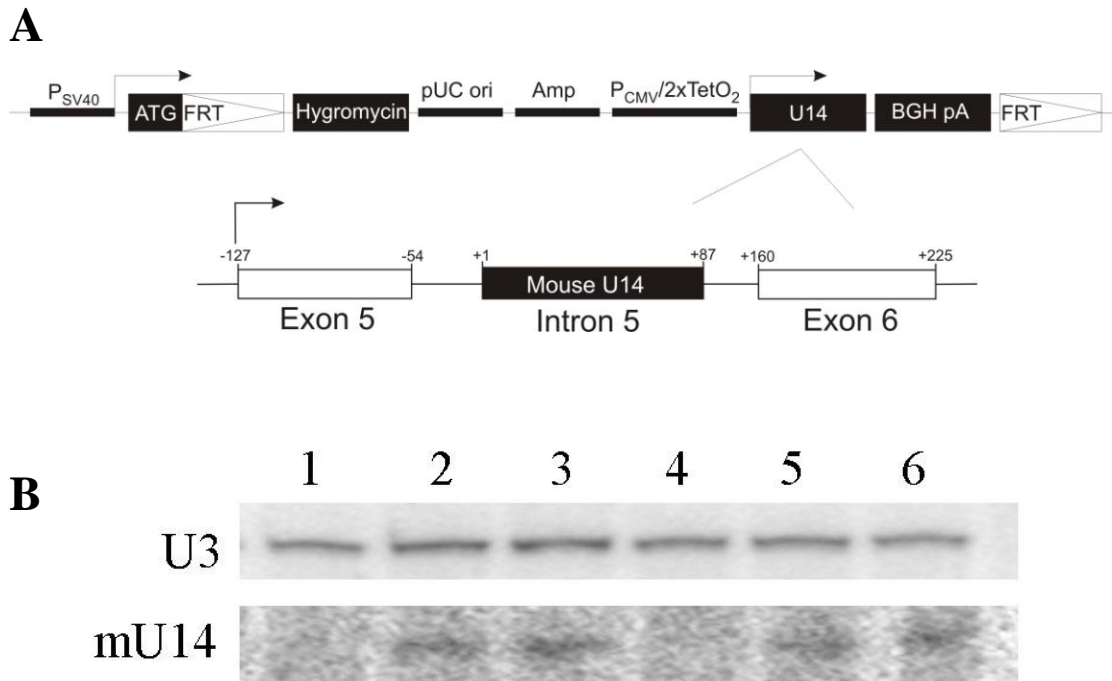


Figure 3.10: Development of a Tet-inducible mU14 snoRNA HEK293 cell line

(A) Schematic diagram of the Tet-inducible mU14 in an Ffp-In T-Rex HEK293 incorporated pcDNA5/FRT/TO expression cassette. FRT marks insertion site. The expression of mouse hsc70, which contained U14, was under the control of the CMV/2xTetO₂ promoter. P_{SV40} represents Simian virus 40 promoter. BGH pA represents Bovine growth hormone polyadenylation signal. For further details of plasmid see Chapter 2.2.1

(B) Primer extension assays showing expression of mU14 snoRNA from the Tet-inducible mU14 HEK293 cell line. The mU14 HEK293 cells were induced by the addition of varying quantities of Tet for different lengths of time. After induction cells were harvested, RNA extracted and primer extension assays performed using primers specific for the mU14 and U3 snoRNAs. The U3 primer extension assay was performed as a control of sample loading. The samples were separated by electrophoresis on 8 % acrylamide gel containing 7 M urea gel and results viewed by autoradiography. Labels left of panel indicate primer extension probes used. Labels above panel, 1 = no Tet; 2 = 2 µg/ml Tet, 48 hours; 3 = 2 µg/ml Tet, 72 hours; 4 = 1 µg/ml Tet, 48 hours; 5 = 1 µg/ml Tet, 72 hours; 6 = 0. 5 µg/ml Tet, 72 hours.

3.2.6. Inducible protein systems

To further study box C/D snoRNP biogenesis systems that enabled the controlled expression of a tagged protein of interest were constructed. These included two Tet-inducible HEK293 cell lines, expressing tagged fibrillarin or SMN, and a reporter plasmid system expressing tagged fibrillarin. These inducible proteins systems provide further methods of analysing box C/D snoRNP biogenesis by determining the localisation and interactions of newly synthesised fibrillarin and SMN.

Tet-inducible tagged fibrillarin and SMN HEK293 cell lines were produced as described earlier in Chapter 3.2.5 and 2.2.3. In order for the induced fibrillarin and SMN to be specifically detected they were expressed with N terminal 2 x FLAG and 6 x His fusion tags (Figure 3.11). A control HEK293 cell line was also developed that only expressed the 2 x FLAG and 6 x His fusion tags, which was utilised as a negative control in some experiments. Before these cells lines were used to analyse box C/D snoRNP biogenesis it was necessary to ensure that the expressed protein behaved in a similar manner to the respective endogenous protein.

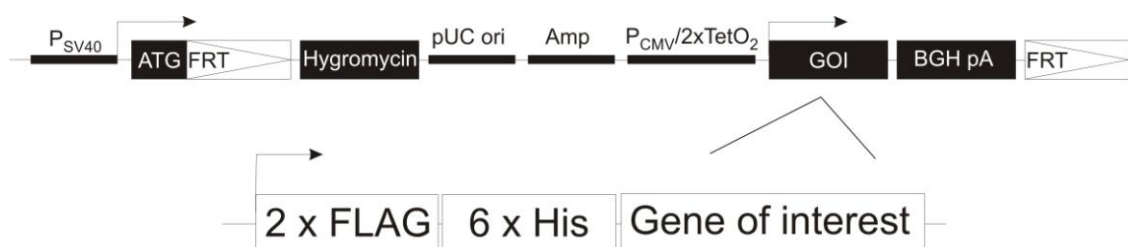


Figure 3.11: Schematic diagram of the Tet-responsive fibrillarin and SMN HEK293 expression cassette

Schematic diagram of the Tet-inducible 2 x FLAG 6 x His tagged SMN and fibrillarin pcDNA5/FRT/TO expression cassette. The FRT recombination site indicates the insertion site of the pcDNA5/FRT/TO expression plasmid into the Flp-In T-Rex HEK293 cells. The expression of SMN or fibrillarin is under the control of the CMV/2xTetO₂ promoter. P_{SV40} represents Simian virus 40 promoter. BGH pA represents Bovine growth hormone polyadenylation signal. For further details of plasmid see Chapter 2.2.3.

To determine the time required for 2 x FLAG 6 x His fibrillarin (FLAG-fibrillarin) expression the FLAG-fibrillarin HEK293 cells were induced by 1 µg/ml of Tet. As a control FLAG-fibrillarin HEK293 cells were grown without the addition of Tet (0 hour). After six, twelve and twenty-four hours cells were harvested, proteins separated on a 12 % SDS-PAGE gel and Western blot assays performed using anti-fibrillarin antibodies. It was possible to distinguish between endogenous and FLAG-fibrillarin using anti-fibrillarin antibodies as the induced FLAG-fibrillarin migrated significantly slower than endogenous fibrillarin on SDS-PAGE gels. A Western blot assay was also performed using anti-nucleolin antibodies to ensure equal sample loading. In the absence of Tet FLAG-fibrillarin was not expressed; however, six hours after the addition of Tet FLAG-fibrillarin was expressed at a slightly lower level than endogenous fibrillarin (Figure 3.12). In cells incubated with Tet for twenty-four hours there was an increase in the levels of FLAG-fibrillarin; however, the levels were still not quite as high as the endogenous fibrillarin. This experiment showed that FLAG-fibrillarin was expressed as early as six hours after the addition of 1 µg/ml of Tet at levels comparable to endogenous fibrillarin. This was also an important control to show that the addition of 1 µg/ml of Tet did not result in over expression of FLAG-fibrillarin.

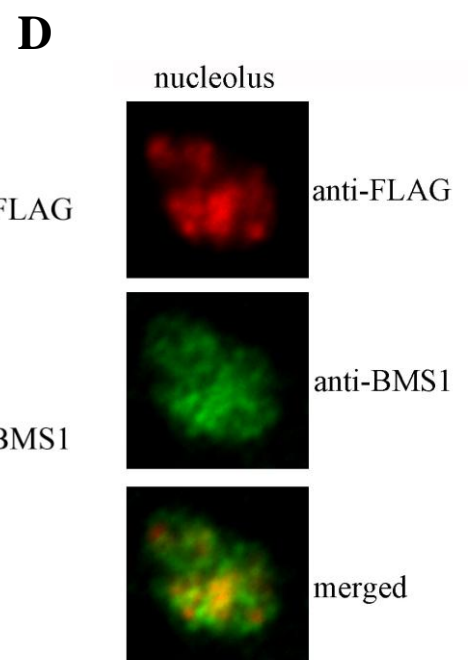
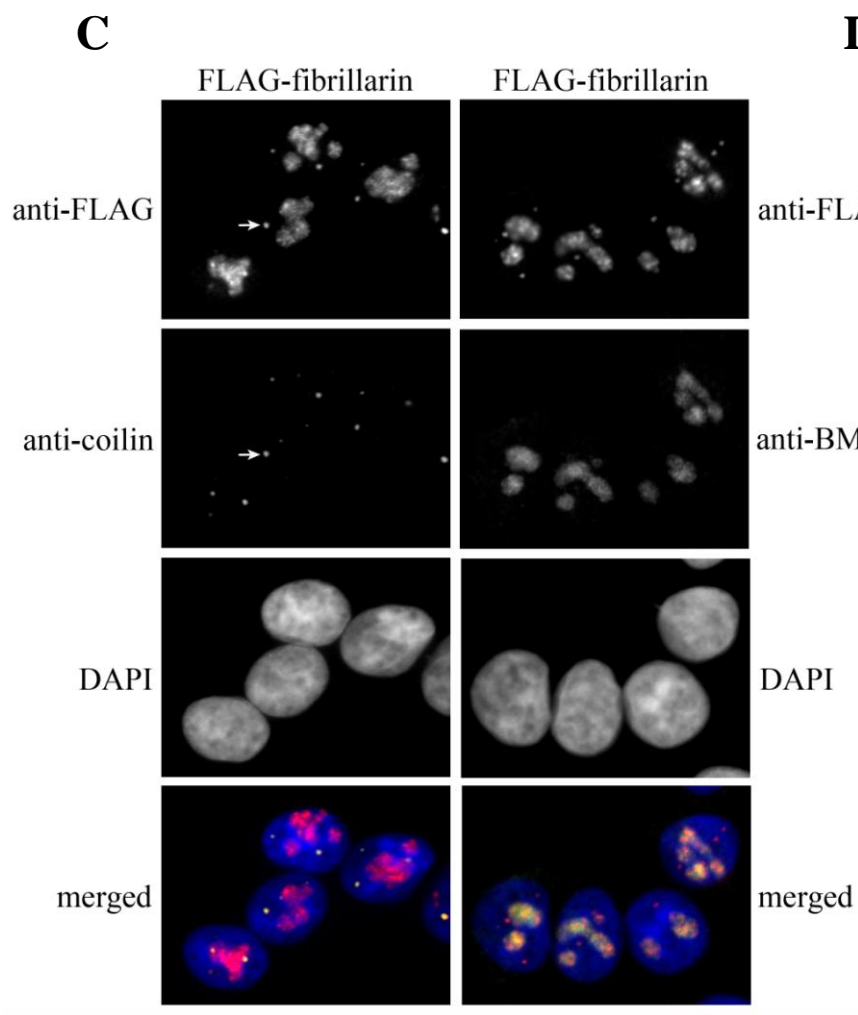
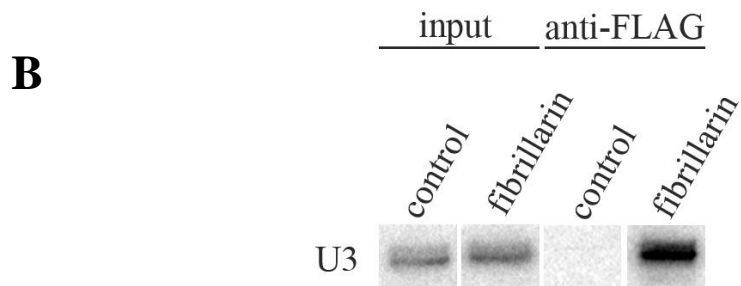
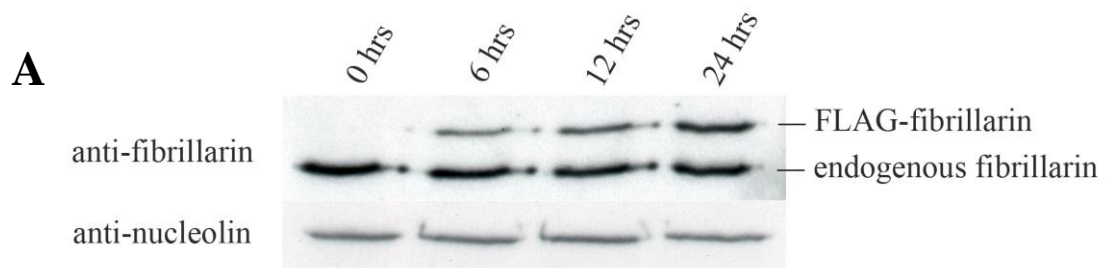
Figure 3.12: Tet-inducible FLAG-fibrillarin HEK293 cell line controls

(A) Western blot assay showing the Tet induction of FLAG-fibrillarin HEK293 cells. FLAG-fibrillarin HEK293 cells were induced by addition of 1 µg/ml of Tet. As a control FLAG-HEK293 cells that had not been induced were harvested (0 hrs). At several time points (indicated above gel) cells were harvested, proteins separated on a 12 % SDS-PAGE gel and Western blot assays performed using anti-nucleolin and anti-fibrillarin antibodies (indicated left of panel). The migration of proteins is shown to the right of the panel.

(B) Co-precipitation of the U3 snoRNA with FLAG-fibrillarin. Whole cell extracts were produced from induced HEK293 cells that expressed FLAG-fibrillarin or just the FLAG tag (control). Immunoprecipitation assays were performed using anti-FLAG antibodies and co-precipitating RNA extracted, separated by electrophoresis on an 8 % acrylamide gel containing 7 M urea and analysed by Northern blot hybridisation using a radiolabelled probe specific for the U3 box C/D snoRNA. Results were visualised by autoradiography. Input represents 10 % of the amount of extract used in assay.

(C) Immunofluorescence of induced FLAG-fibrillarin HEK293 cells. FLAG-fibrillarin cells, grown on coverslips, were induced by the addition of 1 µg/ml of Tet. Forty-eight hours after induction cells were fixed and immunofluorescence performed using anti-FLAG antibodies to identify the localisation of FLAG-fibrillarin. Anti-coilin and anti-BMS1 antibodies were used to indicate the location of Cajal bodies and nucleolus, respectively. DAPI staining of DNA was used to indicate the nucleoplasm. Antibodies and markers used are indicated to the left and right of the panel. The bottom row of images is a merged image of FLAG-fibrillarin (red) with either coilin (green) or BMS1 (green) as indicated and DAPI (blue). The white arrows indicate the location of a Cajal body. The bar to the bottom right of the panel represents 5 µm.

(D) Enlarged nucleolus from Figure 3.12 C. Antibodies used are indicated to the right of the panel. The bottom row is a merged image of FLAG-fibrillarin (red) and BMS1 (green).



To determine whether FLAG-fibrillarin associated with the box C/D snoRNPs the FLAG-fibrillarin HEK293 cells were induced by the addition of 1 µg/ml of Tet. As a negative control 1 µg/ml Tet was also added to the control HEK293 cell line. After forty-eight hours cells were harvested and whole cell extracts produced. To perform immunoprecipitation assays anti-FLAG antibodies were bound to protein A Sepharose beads and incubated with the cell extract of interest. The beads were washed, RNA extraction performed and co-precipitating RNAs separated by electrophoresis on an 8 % acrylamide gel containing 7 M urea. RNA was analysed by Northern blot hybridisation using a radiolabelled probe specific for the U3 box C/D snoRNA and results visualised by autoradiography. Anti-FLAG antibodies successfully co-precipitated U3 snoRNA from FLAG-fibrillarin HEK293 cell extracts but not from the control HEK293 cell extracts (Figure 3.12 B). In comparison to the 10 % input it appears that 20-30 % of the total cell U3 snoRNA was co-precipitated from the FLAG-fibrillarin HEK293 cell extract. This indicates that the FLAG-fibrillarin can efficiently associate with the box C/D snoRNPs.

To determine if the FLAG-fibrillarin localised correctly FLAG-fibrillarin HEK293 cells, grown on cover slips, were induced by the addition of 1 µg/ml Tet. Forty-eight hours later cells were fixed and analysed by immunofluorescence. Anti-FLAG antibodies were used to detect FLAG-fibrillarin. Anti-coilin and anti-BMS1 antibodies were used to determine locations of the Cajal bodies and nucleolus, respectively. BMS1 was distributed throughout the nucleolus, indicating localisation to both the DFC and GC, which was representative of its role in both early and late pre-rRNA processing (Figure 3.12 D; Wegierski et al, 2001). FLAG-fibrillarin was only found in discrete sub-compartments of the nucleolus, indicating it was localised specifically to the DFC. FLAG-fibrillarin also co-localised with coilin, which indicates it was present in the Cajal bodies (Figure 3.12 C). Therefore, FLAG-fibrillarin shows the same localisation profile as endogenous fibrillarin (see Figure 3.1; Tollervey et al, 1991). Based on these controls it appears that induced FLAG-fibrillarin behaves in the same manner as endogenous fibrillarin.

It was also necessary to analyse the Tet-inducible SMN HEK293 cell line. First the level of expression of 2 x FLAG 6 x His tagged SMN (FLAG-SMN) was analysed. To determine the time required for expression FLAG-SMN HEK293 cells were grown in the presence of 1 µg/ml of Tet. As a control FLAG-SMN HEK293 cells that had not been induced were also harvested (0 hour). After six, twelve and twenty-four hours incubation with Tet cells were harvested, proteins separated on a 12 % SDS-PAGE gel and Western blot assays performed using anti-SMN antibodies. It was possible to distinguish between the FLAG-SMN and endogenous SMN using only anti-SMN antibodies as the induced FLAG-SMN migrated significantly slower than endogenous SMN on SDS-PAGE gels. In the absence of Tet no FLAG-SMN was expressed; however, twelve hours after the addition of Tet there was a detectable accumulation of FLAG-SMN which continued to increase up to twenty-four hours (Figure 3.13 A). The level of FLAG-SMN detected after twenty-four hours was similar to the level of endogenous SMN. This also showed that the 1 µg/ml of Tet did not result in the over expression of FLAG-SMN.

As SMN has previously been shown to bind to pre-snRNAs during snRNP biogenesis (Pellizzoni et al, 2002; Wehner et al, 2002; Yong et al, 2002, 2004) an immunoprecipitation assay was performed to determine whether the FLAG-SMN was also able to associate with snRNA. In this experiment FLAG-SMN and control HEK293 cells were induced by the addition of 1 µg/ml of Tet. After forty-eight hours whole cell extracts were produced and immunoprecipitation assays performed using anti-FLAG antibodies as described earlier. Co-precipitating RNAs were separated by electrophoresis on an 8 % acrylamide gel containing 7 M urea, Northern blot hybridisation performed using a radiolabelled probe specific for the U1 snRNA and results visualised using autoradiography. Anti-FLAG antibodies successfully co-precipitated the U1 snRNA from FLAG-SMN HEK293 cell extracts but not from the control HEK293 cell extracts (Figure 3.13 B). In relation to the 10 % input it appears that 2-3 % of the total cell U1 snRNA has been co-precipitated from the FLAG-SMN HEK293 cell extract. The small percentage of total cell snRNA precipitated was expected as SMN only interacts with pre-snRNA and is not stably associated into the snRNP. This experiment showed that the FLAG-SMN was able to associate with snRNA, which is characteristic of endogenous SMN.

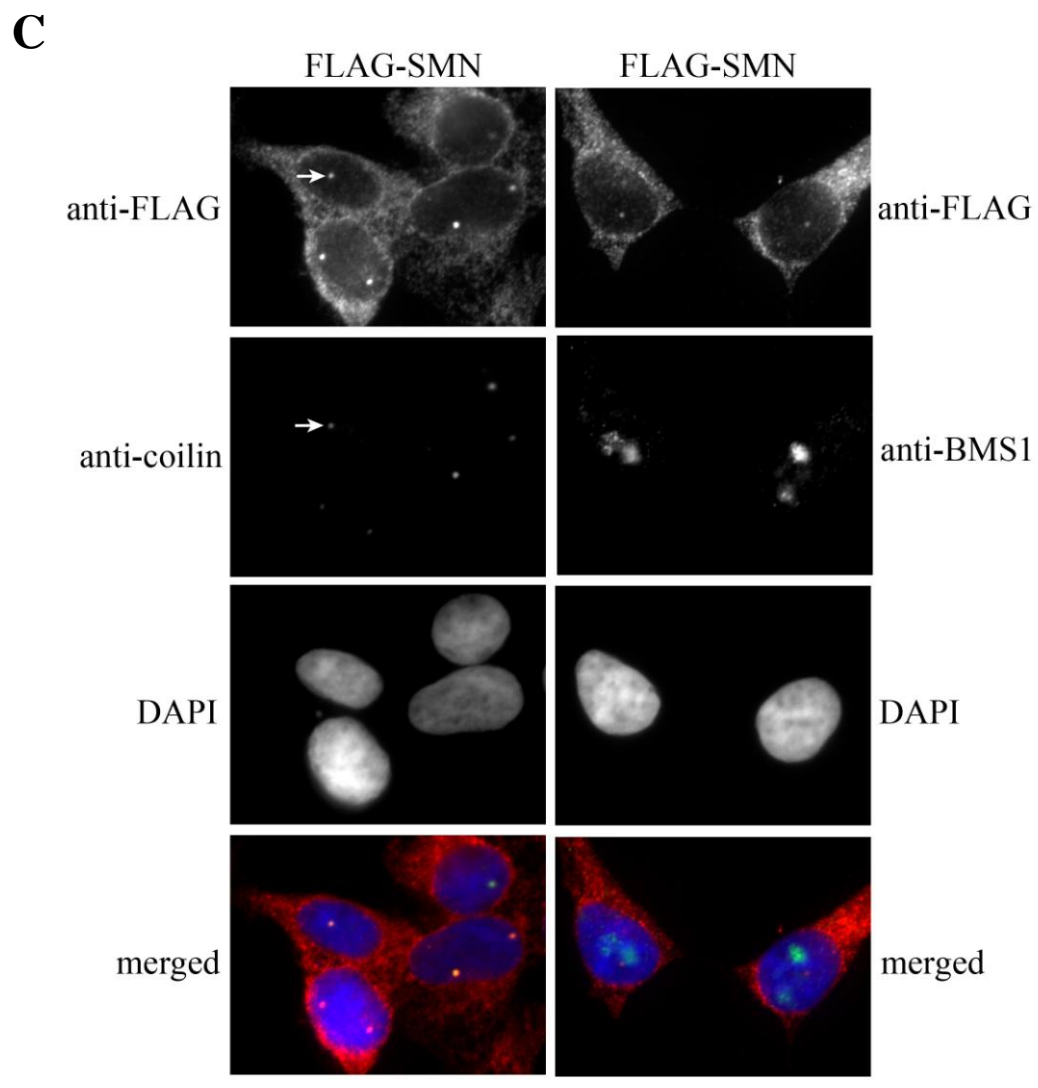
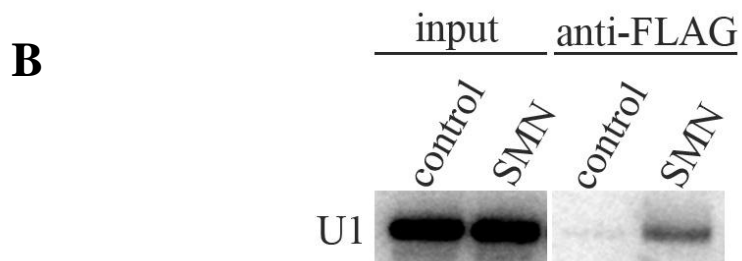
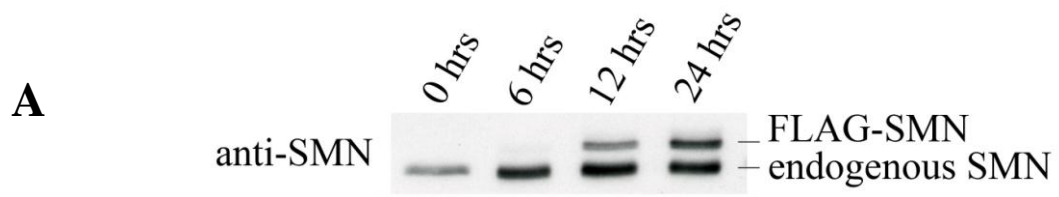
Immunofluorescence was performed to determine whether the localisation of FLAG-SMN was the same as endogenous SMN. FLAG-SMN HEK293 cells, grown on cover slips, were induced by 1 µg/ml of Tet. After forty-eight hours cells were fixed and immunofluorescence performed using anti-FLAG antibodies to determine the localisation of FLAG-SMN. Anti-coilin and anti-BMS1 antibodies were used to detect the Cajal bodies and nucleolus, respectively. FLAG-SMN was found in the cytoplasm and enriched in Cajal bodies, which is characteristic of endogenous SMN (Liu and Dreyfuss, 1996). Taken together all the data indicates that the FLAG-SMN was comparable to endogenous SMN.

Figure 3.13: Tet-inducible FLAG-SMN HEK293 cell line controls

(A) Western blot assay showing Tet induction of FLAG-SMN HEK293 cells. FLAG-SMN HEK293 cells were induced by addition of 1 µg/ml of Tet. As a control FLAG-HEK293 cells that had not been induced were harvested (0 hrs). At several time points cells were harvested, proteins separated on an 12 % SDS-PAGE gel and Western blot assays performed using anti-SMN antibodies to detect both FLAG-SMN and endogenous SMN (indicated left of panel). The time periods are shown above panel. The migration of proteins is shown to the right of the panel.

(B) Co-precipitation of U1 snRNA with FLAG-SMN. Whole cell extracts were produced from induced HEK293 cells that expressed FLAG-SMN or just the FLAG tag (control). Immunoprecipitation assays were performed using anti-FLAG antibodies and co-precipitating RNA extracted, separated by electrophoresis on an 8 % acrylamide gel containing 7 M urea and analysed by Northern blot hybridisation using a probe specific for the U1 snRNA. Results were visualised by autoradiography. Input represents 10 % of the amount of extract used in assay.

(C) Immunofluorescence of induced FLAG-SMN HEK293 cells. FLAG-SMN cells, grown on coverslips, were induced by the addition of 1 µg/ml of Tet. Forty-eight hours after induction cells were fixed and immunofluorescence performed using anti-FLAG antibodies to identify the localisation of FLAG-SMN. Anti-coilin and anti-BMS1 antibodies were used to indicate the location of Cajal bodies and nucleolus, respectively. DAPI staining of DNA was used to indicate the nucleoplasm. Antibodies and markers used indicated left and right of panel. The bottom row is a merged image of FLAG-SMN (red) with either coilin (green) or BMS1 (green) and DAPI (blue). The White arrows indicate the location of a Cajal body. The bar shown to the bottom right of the panel represents 5 µm.



One of the experiments the Tet-inducible FLAG-SMN HEK293 cell line was created to be used in was siRNA rescue assays. In these experiments endogenous SMN was to be depleted by siRNAs and the effect of the loss of SMN on box C/D snoRNP biogenesis rescued by induction of FLAG-SMN expression. In order to maintain the levels of the induced FLAG-SMN siRNAs were designed against the 3'UTR of endogenous SMN mRNA, which was not present in the FLAG-SMN mRNA. Even though the siRNA target sequence was not present in the FLAG-SMN mRNA it was still necessary to determine, experimentally, whether induced FLAG-SMN was able to accumulate in cells transfected with the siRNAs targeting SMN. For this control FLAG-SMN HEK293 cells were transfected with siRNAs targeting either firefly luciferase (control) or SMN and simultaneously induced by the addition of 1 μ g/ml of Tet. Cells were harvested after twenty-four, forty-eight and seventy-two hours. Proteins were separated on a 12 % SDS-PAGE gel and Western blot assays performed using anti-SMN antibodies to identify both FLAG-SMN and endogenous SMN as described earlier. A Western blot was also performed using anti-nucleolin antibodies as a control of sample loading.

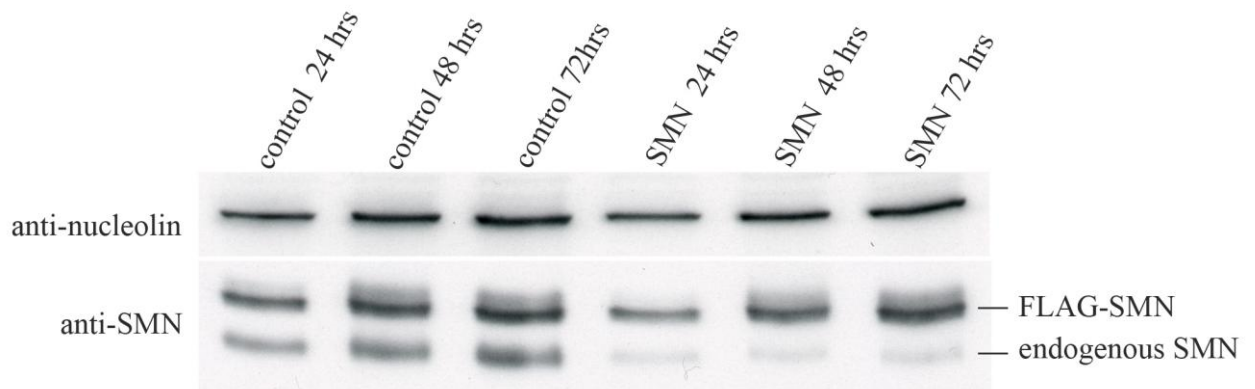


Figure 3.14: Depletion of endogenous SMN in induced FLAG-SMN HEK293 cells
 FLAG-SMN HEK293 cells were transfected with siRNAs targeting either firefly luciferase (control) or SMN and induced by the addition of 1 μ g/ml of Tet. Cells were harvested after twenty-two, forty-eight and seventy-two hours, proteins separated on a 12 % SDS-PAGE gel and Western blot assays were performed using anti-SMN antibodies to identify both endogenous SMN and FLAG-SMN. A Western blot was also performed using anti-nucleolin antibodies as a control of sample loading. The protein targeted for depletion is indicated above the panel. Antibodies used in the Western blot are shown to the left of the panel. The migration of proteins is shown to the right of the panel.

In cells transfected with siRNAs targeting SMN mRNA there was a reduction in the level of endogenous SMN, compared to the control cells, after as little as twenty-four hours (Figure 3.14). The level of the induced FLAG-SMN was unaffected in cells transfected with siRNAs targeting SMN, which indicates that the SMN siRNAs are unable to target the FLAG-SMN mRNA for degradation.

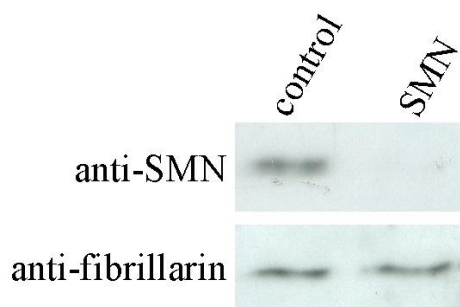


Figure 3.15: Depletion of SMN in HeLa SS6 cells

HeLa SS6 cells were transfected with siRNAs targeting firefly luciferase (control) or SMN. After sixty hours cells were harvested, proteins separated on a 12 % SDS-PAGE gel and Western blot assays performed using antibodies that recognised SMN. A Western blot assay was also performed using anti-fibrillarin antibodies as a control of sample loading. The protein targeted for depletion is indicated above the panel. The antibodies used are indicated to the left of the panel.

The depletion of endogenous SMN in the FLAG-SMN HEK293 cells (Figure 3.14) was not as efficient as was seen when the same siRNA duplex was used to deplete SMN in HeLa SS6 cells (Figure 3.15). To determine whether it was possible to use siRNAs to efficiently deplete target proteins in HEK293 cells siRNAs targeting NOP56, NOP58 and fibrillarin were transfected into basic HEK293 cells and the localisation of the U3 box C/D snoRNA was analysed by FISH as previously described (Chapter 3.2.1). In HeLa SS6 cells the depletion of NOP58 resulted in the dramatic reduction in the level of the U3 snoRNA while the depletion of NOP56 and fibrillarin resulted in the accumulation of the U3 snoRNA in the nucleoplasm and Cajal bodies (Figure 3.4). However, in HEK293 cells transfected with siRNAs targeting NOP56, NOP58 and fibrillarin there was no change in the levels or distribution of the U3 box C/D snoRNA (data not shown). A possible explanation for the lack of effect on the U3 snoRNA levels and localisation in HEK293 cells is that the siRNA mediated depletion was not as efficient in HEK293 cells as in HeLa SS6 cells. To try and overcome this issue both chemical and electroporation transfection protocols were employed; however, these

proved equally unsuccessful. Due to these issues the use of the inducible HEK293 cells in experiments, which incorporated siRNA depletion, was not possible.

As siRNA mediated depletion proved difficult in HEK293 cells an inducible fibrillarin reporter plasmid was designed for use in HeLa SS6 cells. This system used a pPTuner reporter plasmid (Clontech) containing a degradation domain (DD) tag that when expressed targets the protein, to which it is fused, for proteosomal-mediated degradation. However, when the stabilising agent Shield1 is added the DD is no longer targeted to the proteosome and the protein, it is fused with, accumulates (Banaszynski et al, 2006). To create this system fibrillarin cDNA with N terminal 2 x FLAG 6 x His sequences was cloned into the pPTuner plasmid. The pPTuner plasmid (Figure 3.16) also contained an internal ribosome entry site 2 (IRES2) element upstream of a GFP coding sequence that when expressed provides a marker for transfected cells.

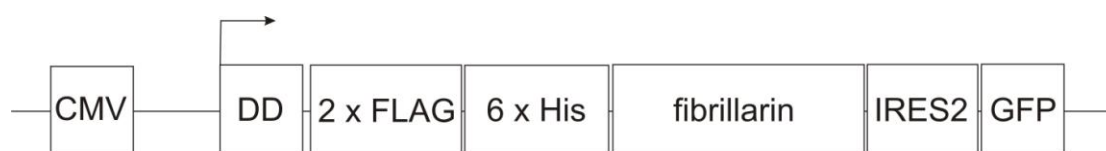


Figure 3.16: Organisation of DD-fibrillarin plasmid

Fibrillarin cDNA with 2 x FLAG and 6 x His N terminal tags was cloned into a pPTuner IRES2 GFP plasmid (Clontech). The pPTuner plasmid contains a CMV (cytomegalo virus) promoter, N terminal degradation domain (DD), internal ribosome entry site 2 (IRES2) and a green florescent protein (GFP) tag. Arrow indicates transcription start site.

The DD fibrillarin reporter plasmid (DD-fibrillarin) was developed to be co-transfected into HeLa SS6 cells with siRNAs. Once target protein levels had been depleted the addition of shield1 would result in the accumulation of DD-fibrillarin and the incorporation of DD-fibrillarin into newly synthesized snoRNPs and its localisation could be analysed. The use of this procedure removed the requirement for multiple transfections as only a single transfection of both siRNA and reporter plasmid was required.

Before using the DD-fibrillarin reporter plasmid to analyse box C/D snoRNP biogenesis it was necessary to first characterise its expression and accumulation. To determine the

quantity of shield1 required to induce stabilisation HeLa SS6 cells were transfected with the DD-fibrillarin reporter plasmid and incubated with various levels of shield1. After twenty-four hours cells were harvested, proteins separated on a 12 % SDS-PAGE and Western blot assays performed using anti-FLAG antibodies to detect DD-fibrillarin. In addition a Western blot assay was performed using anti-fibrillarin antibodies as a control of sample loading. These assays showed that in the absence of shield1 there was some accumulation of DD-fibrillarin; however, when 50 nM to 1000 nM of shield1 was added there was a dramatic accumulation of DD-fibrillarin (Figure 3.17 A). Interestingly, there was very little difference in the accumulation of DD-fibrillarin when 50 nM or 1000 nM of shield1 was added.

To determine the relative level of DD-fibrillarin accumulation in comparison to endogenous fibrillarin further analysis of the Western blot in Figure 3.17 A was performed. Using only anti-fibrillarin antibodies it was possible to distinguish between DD-fibrillarin and endogenous fibrillarin as the DD-fibrillarin migrated slower on SDS-PAGE gels. In the absence of shield1 there was a low level accumulation of DD-fibrillarin (Figure 3.17 B). When 250 nM shield1 was added there was a higher accumulation of DD-fibrillarin; however, the level of DD-fibrillarin was still much lower than the level of endogenous fibrillarin. This was an important control that showed DD-fibrillarin was not over expressed.

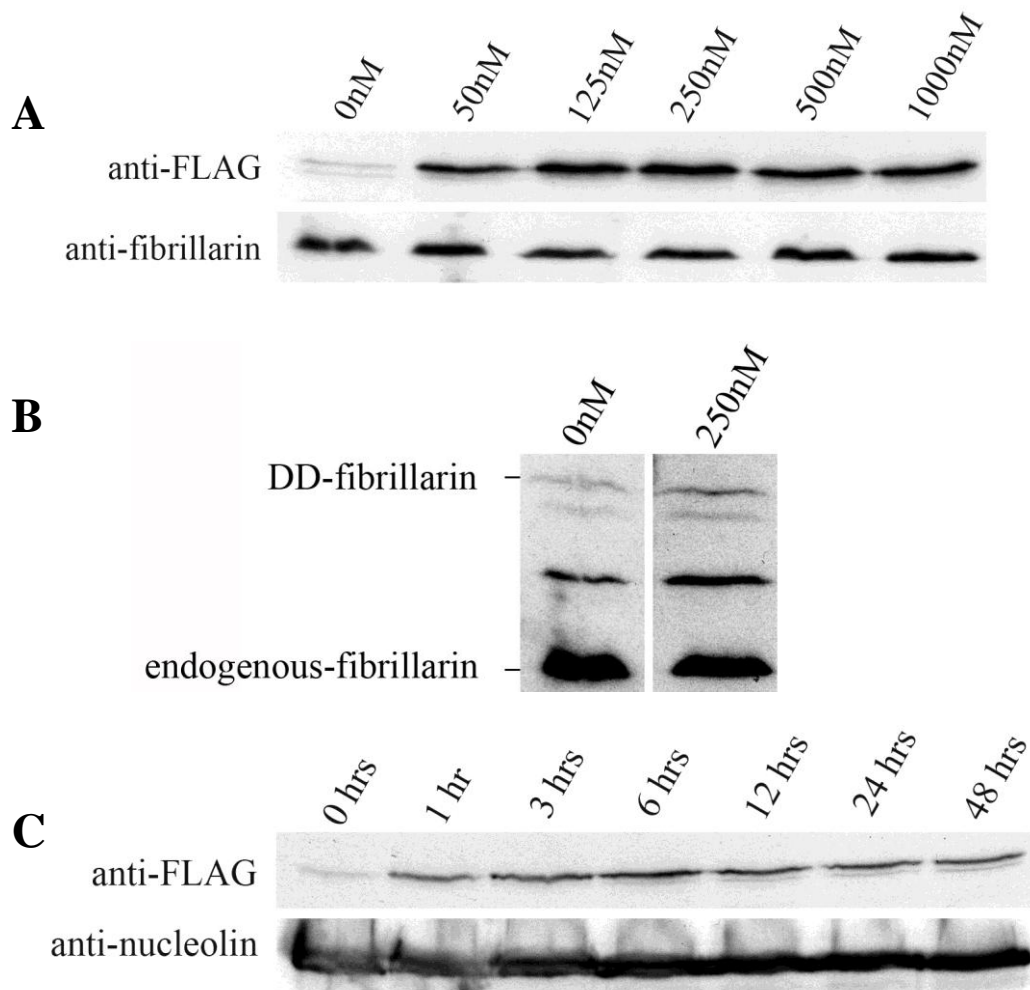


Figure 3.17: Control experiments of DD-fibrillarin accumulation

(A) Western blot assay showing quantity of shield1 required for stabilisation and accumulation of DD-fibrillarin. HeLa SS6 cells were transfected with the DD-fibrillarin reporter plasmid and incubated with various concentrations of shield1 for twenty-four hours. Proteins were separated on a 12 % SDS-PAGE and analysed by Western blot assays using anti-FLAG antibodies. A Western blot assay was also performed using anti-fibrillarin antibodies as a control of sample loading. The quantity of shield1 used is indicated above the panel. The antibodies used in the Western blot are shown to the left of the panel.

(B) Western blot showing the level of DD-fibrillarin accumulation in comparison to endogenous fibrillarin. Longer exposure and larger section of the Western blot performed in Figure 3.17 A with both DD-fibrillarin and endogenous fibrillarin visible using only anti-fibrillarin antibodies (indicated left of panel). The quantity of shield1 used is indicated above the panel.

(C) Western blot showing time course of DD-fibrillarin stabilisation and accumulation. HeLa SS6 cells were transfected with the DD-fibrillarin reporter plasmid and incubated with 250 nM of shield1. As a control cells transfected with the DD-fibrillarin plasmid were extracted before the addition of shield1 (0 hours). Cells were extracted after one, three, six, twelve, twenty-four and forty-eight hours, proteins separated on a 12 % SDS-PAGE and analysed by Western blot assays using anti-FLAG antibodies to identify DD-fibrillarin. A Western blot assay was also performed using anti-nucleolin antibodies to control sample loading. The incubation time is indicated above the panel. The antibodies used in the Western blot are shown to the left of the panel.

A time course of DD-fibrillarin accumulation in the presence of shield1 was also performed. HeLa SS6 cells were transfected with the DD-fibrillarin reporter plasmid and twelve hours later shield1 was added to a final concentration of 250 nM. This concentration of shield1 was selected as it has previously shown to produce a clear accumulation of DD-fibrillarin (Figure 3.17 A). As a control cells transfected with the DD-fibrillarin reporter plasmid were harvested before the addition of shield1 (0 hours). Cells supplemented with shield1 were harvested after one, three, six, twelve, twenty-four and forty-eight hours. Proteins were separated on a 12 % SDS-PAGE gel and Western blot assays performed using anti-FLAG antibodies to detect DD-fibrillarin. As a control of sample loading a Western blot assay was also performed with anti-nucleolin antibodies. In the absence of shield1 there was a slight accumulation of DD-fibrillarin; however, as little as one hour after the addition of shield1 there was a much larger accumulation of DD-fibrillarin (Figure 3.17 C). The level of DD-fibrillarin accumulation did not appear to change after one hour and even after forty-eight hours the level of DD-fibrillarin remained approximately the same.

Immunofluorescence was performed to determine whether the DD-fibrillarin localised the same as endogenous fibrillarin. Two sets of HeLa SS6 cells were transfected with the DD-fibrillarin reporter plasmid and to one set 250 nM of shield1 was added and to the control cells no shield1 was added. Both sets of cells were then incubated for another six hours, fixed and immunofluorescence performed using anti-FLAG antibodies to detect DD-fibrillarin. Anti-coilin and anti-BMS1 antibodies were used to detect the location of the Cajal bodies and nucleolus, respectively.

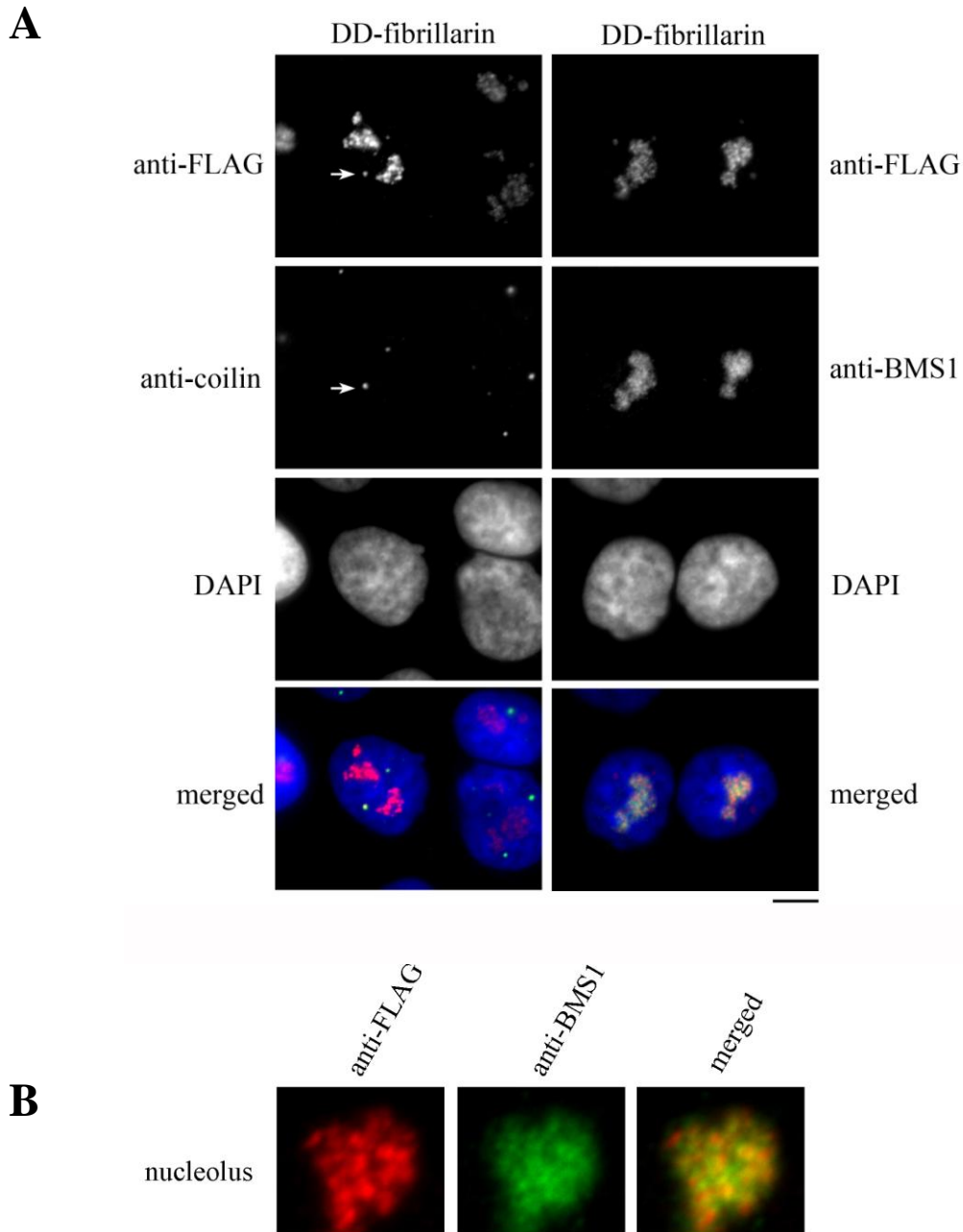


Figure 3.18: Immunofluorescence of DD-fibrillarin

(A) HeLa SS6 cells, grown on coverslips, were transfected with the DD-fibrillarin reporter plasmid. Twelve hours later shield1 was added to a concentration of 250 nM. Six hours after the induction of stabilisation cells were fixed and immunofluorescence performed. Anti-FLAG antibodies were used to detect DD-fibrillarin. Anti-coilin and anti-BMS1 antibodies were used to indicate the locations of the Cajal bodies and the nucleolus, respectively. DAPI staining of DNA was used to determine the location of the nucleoplasm. Antibodies used indicated left and right of the panel. The bottom row is a merged image of DD-fibrillarin (red), Coilin or BMS1 (green) and DAPI (blue). The white arrows indicate a Cajal body. The bar shown to the bottom right of the panel represents 5 μ m.

(B) Enlarged nucleolus from Figure 3.18 A. The antibodies used are indicated above panels. The right hand panel is a merged image of FLAG-fibrillarin (red) and BMS1 (green).

In the cells where shield1 was added DD-fibrillarin localised to the nucleolus and the Cajal bodies as was shown by the co-localisation with BMS1 and coilin, respectively (Figure 3.18). BMS1 was distributed throughout the nucleolus in both the DFC and GC, which was representative of its role in both early and late pre-rRNA processing (Wegierski et al, 2001). DD-fibrillarin was found in discrete sub-compartments of the nucleolus, which indicated it was localised specifically to the DFC where the majority of box C/D snoRNPs are found (Figure 3.18 B; see Figure 3.1 for comparison of nucleolar localisation). This experiment showed that DD-fibrillarin localised to the same sub-cellular domains as endogenous fibrillarin (Tollervey et al, 1993). In the control cells a low level background of DD-fibrillarin was visible but the signal was too low to obtain images (data not shown). The results were therefore in line with the Western blot analysis with cells not supplemented with shield1 producing a low level background of DD-fibrillarin but an accumulation of DD-fibrillarin once shield1 was added (Figure 3.17).

One major issue that became apparent from initial experiments with the DD-fibrillarin reporter plasmid was that the GFP marker, translated from the IRES2 element, was not expressed at high enough levels to provide an indication of the level of transfection. The cause of the low GFP expression was due to the fact that proteins encoded after the IRES2 element are translated at a rate three to four times lower than the gene of interest (personal communication from Clontech). To try and amplify the GFP signal immunofluorescence was performed using rabbit anti-GFP antibodies in conjunction with GFP conjugated anti-rabbit antibodies; however, still no GFP signal was visible. The GFP marker was a vital control for experiments utilising siRNA mediated depletion experiments to ensure that when analysing the localisation of DD-fibrillarin, cells with a similar level of transfection could be selected.

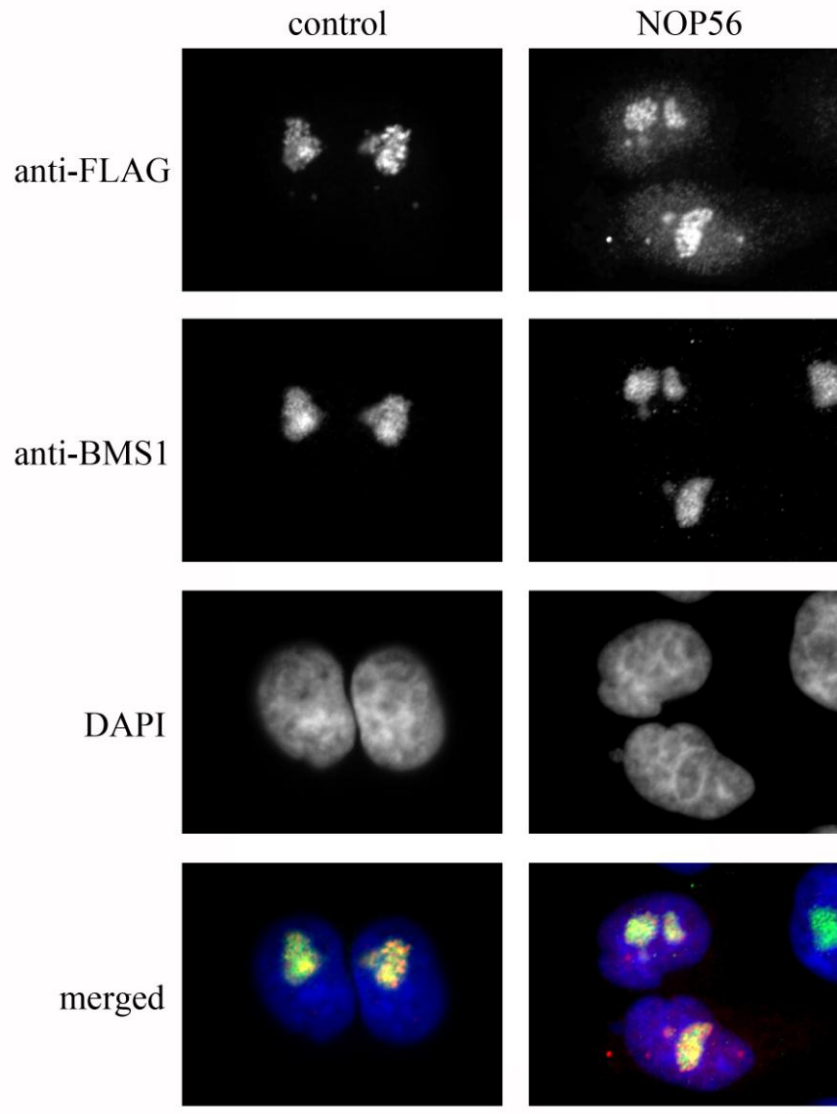


Figure 3.19: The localisation of DD-fibrillarin in NOP56 depleted cells

HeLa SS6 cells were transfected with the DD-fibrillarin reporter plasmid and siRNAs targeting either firefly luciferase (control) or NOP56. Forty-eight hours later the stabilising agent, shield1, was added to a concentration of 250 nM. After six hours cells were fixed and immunofluorescence performed. Anti-FLAG antibodies were used to detect DD-fibrillarin and anti-BMS1 antibodies used to identify the nucleolus. DAPI staining of the DNA was used to determine the location of the nucleoplasm. The same exposure time was used for each probe to allow direct comparison of protein distribution and levels. The protein targeted for depletion is indicated at the top of each column of images. The antibodies used are indicated to the left of the panel. The bottom row is a merged image of DD-fibrillarin (red), BMS1 (green) and DAPI (blue). The bar to the bottom right of the panel represents 5 μ m.

To determine whether the co-transfection of siRNAs and the DD-fibrillarin reporter plasmid would work HeLa SS6 cells were co-transfected with the DD-fibrillarin reporter plasmid and siRNAs targeting firefly luciferase (control) or NOP56. Forty-eight hours later shield1 was added to a final concentration of 250 nM. After another six hours the HeLa SS6 cells were fixed and analysed by immunofluorescence using anti-FLAG antibodies to detect DD-fibrillarin and anti-BMS1 antibodies to identify the nucleolus.

In the control cells DD-fibrillarin was found in the nucleolus, as was shown by the co-localisation with the BMS1 (Figure 3.19). In cells depleted of NOP56 DD-fibrillarin was still found in the nucleolus; however, DD-fibrillarin also accumulated in the nucleoplasm. The accumulation in the nucleoplasm could be due the formation of pre-snoRNP complexes that contained DD-fibrillarin but lacked NOP56 and therefore were unable to localise to the nucleolus.

The co-transfection of the DD-fibrillarin reporter plasmid with siRNAs was only successful using siRNAs targeting NOP56, and even with this depletion only a few cells expressed a detectable amount of DD-fibrillarin. A number of siRNAs targeting other snoRNP factors were tested; however, these resulted in the complete lack of DD-fibrillarin accumulation. A possible explanation for the lack of DD-fibrillarin expression may be that in the absence of the box C/D snoRNP factors DD-fibrillarin was turned over rapidly. Due to an inability to obtain detectable expression of the GFP transfection marker and the issues with DD-fibrillarin accumulation in cells depleted of factors of interest the use of the DD-fibrillarin reporter plasmid was limited.

3.3. Discussion

In this chapter a variety of assays, procedures and approaches were developed to investigate box C/D snoRNP biogenesis. These included inducible cell lines, procedures to analyse *de novo* snoRNP biogenesis and the use of more basic assays including FISH and immunofluorescence to determine the localisation of the box C/D snoRNA, common core proteins and assembly factors.

Box C/D snoRNAs are transcribed in the nucleoplasm and formation of the snoRNP has been predicted to occur in the nucleoplasm, Cajal bodies and cytoplasm (Terns et al, 1995; Boulon et al, 2004; Watkins et al, 2004, 2007; Lemm et al, 2006; McKeegan et al, 2007). One approach used to shed more light on the site(s) of box C/D snoRNP biogenesis was to determine the cellular localisation of the box C/D snoRNP assembly factors. Immunofluorescence analysis revealed that the box C/D snoRNP assembly factors TIP48, TIP49, BCD1, TAF9 and NUFIP localise throughout the nucleoplasm and are not enriched in any nuclear body (Figure 3.6 A and B). The nucleoplasmic localisation of the assembly factors fits with box C/D snoRNP biogenesis occurring in the nucleoplasm (Terns et al, 1995; Boulon et al, 2004; Watkins et al, 2004). As no assembly factors were found enriched in the Cajal bodies this suggests that the assembly factors do not function in box C/D snoRNP biogenesis at this site. It was not, however, possible to entirely rule out the presence of assembly factors in the Cajal bodies.

As many of the box C/D snoRNP assembly factors are involved in other processes in the nucleoplasm it is difficult to draw a firm conclusion on the site(s) of box C/D snoRNP biogenesis by analysing their localisation alone. For instance TAF9 is a TATA binding protein and therefore functions in RNA polymerase II transcription in the nucleoplasm (Milgrom et al, 2005). NUFIP is a zinc finger transcription factor and so is also localised to the nucleoplasm (Cabart et al, 2004). TIP48 and TIP49 have roles in chromatin-modifying complexes (Shen et al, 2000) and transcription activating complexes (Cho et al, 2001) that require these proteins to be localised again to the nucleoplasm. No additional functions outside of box C/D snoRNP biogenesis have been found for BCD1 and therefore, in this case, the nucleoplasmic localisation may solely reflect its role in box C/D snoRNP biogenesis.

Another approach used to analyse box C/D snoRNP biogenesis was the use of siRNA mediated depletion followed by analysing the localisation of box C/D snoRNA. To establish this approach siRNAs were used to deplete the box C/D snoRNP common core proteins NOP56, NOP58 and fibrillarin and FISH used to determine the localisation of the U3 box C/D snoRNA. In the control cells the U3 snoRNA localised predominately to the nucleolus and was also found in the Cajal bodies (Figure 3.4), the typical distribution of U3 snoRNA (Granneman et al, 2004). In cells depleted of NOP56 or fibrillarin there was a slight reduction in the levels of nucleolar U3 snoRNA, more than likely, due to a reduction in the rate of box C/D snoRNP biogenesis. In NOP56 and fibrillarin depleted cells U3 snoRNA accumulated in the nucleoplasm and Cajal bodies, providing further evidence that box C/D snoRNP biogenesis occurs at these sites. The accumulation of box C/D snoRNA in the nucleoplasm when fibrillarin and NOP56 were depleted is in agreement with studies in yeast (Verheggen et al, 2001). A possible explanation is that as only mature box C/D snoRNPs, containing all the common core proteins, are predicted to localise to the nucleolus (Verheegen et al, 2001), pre-snoRNPs lacking either fibrillarin or NOP56 accumulated in the nucleoplasm and Cajal bodies. In cells depleted of NOP58 there was a dramatic reduction in nucleolar U3 snoRNA levels with no accumulation in the nucleoplasm or Cajal bodies (Figure 3.4). The dramatic drop in the U3 snoRNA levels indicates that NOP58 is essential for the stability of the snoRNA, which is agreement with previous studies (Lafontaine and Tollervey, 2000; Watkins et al, 2004). As distinctive phenotypes were observed by the siRNA mediated depletion of the common core proteins this indicates that this procedure could be use to analyse the function of other proteins in box C/D snoRNP biogenesis.

One issue with using siRNAs to deplete the common core proteins followed by analysis of the localisation of endogenous box C/D snoRNA was that it was not possible to discriminate between snoRNAs synthesised prior to protein depletion and newly formed snoRNAs. This is an issue when analysing the box C/D snoRNPs as these are stable complexes, which have been predicted to have long half-lives (personal communication from N.J.Watkins, Newcastle University, UK). Similar issues had to be addressed in studies analysing snRNP biogenesis as these complexes are also very stable and have half-lives of between 60 and 88 hours (Fury and Zieve, 1996). A study by Shparagel and Matera 2005 addressed this issue by developing a number of procedures to

specifically monitor the *de novo* biogenesis of the snRNPs. These included the use of an *in vitro* assembly assay and a two-step transfection procedure whereby cells were first transfected with siRNAs followed by a secondary transfection of a reporter plasmid, once target protein levels had been depleted. A variety of approaches were therefore developed in this study to analyse the *de novo* biogenesis of box C/D snoRNPs.

An assembly assay was developed to analyse box C/D snoRNP biogenesis based upon a snRNP assembly assay (Shparagel and Matera, 2005). In this assay an active cell extract, produced from HeLa SS6 cells that had been depleted of a protein of interest using siRNAs, was incubated with radiolabelled U14 box C/D snoRNA. To assess protein association onto the radiolabelled U14 snoRNA immunoprecipitation assays were performed using antibodies that recognised the box C/D snoRNP common core proteins. Unfortunately, despite the use of a variety of cell extraction protocols it was not possible to immunoprecipitate any newly formed box C/D snoRNPs, even using HeLa SS6 cell extract that had not been subjected to siRNA mediated depletion.

A two-step transfection procedure was developed to analyse *de novo* box C/D snoRNP biogenesis in this study. This involved HeLa SS6 cells being first transfected with siRNAs followed by the secondary transfection of a plasmid that expressed tagged U3 snoRNA (StreptoTag U3; Figure 3.8). In cells depleted of NOP56 the newly synthesised StreptoTag U3 snoRNA accumulated in the nucleolus, nucleoplasm and Cajal bodies, which further implicates the nucleoplasm and Cajal bodies as sites of box C/D snoRNP biogenesis. In cells depleted of NOP58 there was no accumulation of the StreptoTag U3 snoRNA in the nucleoplasm or Cajal bodies. Interestingly, even with this *de novo* procedure the StreptoTag U3 snoRNA still accumulated in the nucleolus of cells depleted of NOP56 and NOP58, which indicate that a low level of box C/D snoRNP biogenesis was still occurring. An issue with this assay is that without an internal marker of transfection, such as a GFP marker co-expressed from the plasmid, it is difficult to compare the relative levels of the StreptoTag U3 snoRNA between cells making analysis problematic.

A variety of other siRNAs targeting fibrillarin and box C/D snoRNP assembly factors were tested using this two-step transfection procedure; however, these experiments were not successful. The main issue that limited the use of this system was the poor rate of

secondary transfection of the StreptoTag U3 reporter plasmid. A number of different transfection protocols were utilised to try and overcome this issue but these proved unsuccessful. While it was not possible to use this procedure to further study box C/D snoRNP biogenesis it provided clarification that the results seen with the endogenous U3 snoRNA after NOP56 and NOP58 depletion are accurate (Figure 3.4).

To remove the need for multiple transfections Tet-inducible box C/D snoRNA HEK293 cell lines were created that expressed the StreptoTag U3 or mU14 snoRNAs. These cell lines allowed a single transfection of siRNAs to be performed followed by induction of box C/D snoRNA once target protein levels have been depleted. Both these cell lines expressed the respective snoRNA by the addition of Tet and could be specifically detected above endogenous snoRNA. However, despite the use of varying quantities of Tet and numerous incubation times neither of the box C/D snoRNA HEK293 cell lines expressed sufficient levels of the respective snoRNA to be comfortably detected, therefore the use of these cell lines was limited. One possible explanation for the low level of the StreptoTag U3 and mU14 snoRNA expression from the Tet-inducible HEK293 cells is that in nature multiple copies of these RNAs are present in the genome, which could reflect that each copy is only expressed at low levels. For example in rat, human and mouse hsc70, three copies of the U14 snoRNA are expressed from introns 5, 6 and 8 (Sorgner and Pelhama, 1987; Dworniczak and Mirault, 1987; Lui and Maxwell, 1990). Furthermore, two copies of the U14 snoRNA are expressed from ribosomal protein genes (Kenmochi et al, 1996). In regards the U3 snoRNA there are four to six copies in the human genome (Mazan et al, 1993). One potential method of increasing the expression levels of the inducible snoRNAs is to insert tandem repeats of the snoRNA genes into the Tet-inducible HEK293 expression cassettes. Tandem repeats of snoRNA genes was used successfully by Darzacq et al 2008 to create an E3 box H/ACA inducible U2OS stable cell line that expressed inducible E3 snoRNA at a level comparable with endogenous E3. As the U3 snoRNA is encoded as an independent gene tandem repeats of the entire U3 gene could be inserted into the expression cassette. As mU14 snoRNA is an intronic RNA to increase its expression tandem repeats of the hsc70 gene incorporating U14 i.e. intron 5 with sections of the surrounding exons, would need to be inserted into the HEK293 expression cassette.

Two Tet-inducible tagged protein cell lines were also created to study box C/D snoRNP biogenesis, the FLAG-SMN and FLAG-fibrillarin HEK293 cell lines. These cell lines expressed their tagged proteins at levels that were similar to the respective endogenous proteins (Figure 3.12 and 3.13). Immunofluorescence showed that FLAG-SMN localised to the cytoplasm and Cajal bodies, and FLAG-fibrillarin localised to the nucleolus and Cajal bodies (Figure 3.1). Immunoprecipitation assays using anti-FLAG antibodies confirmed that FLAG-SMN was able to bind snRNA and the FLAG-fibrillarin could associate with box C/D snoRNPs (Figures 3.12 and 3.13). This indicates that both FLAG-SMN and FLAG-fibrillarin function in a similar manner to their respective endogenous proteins (Tollervey et al, 1993; Liu and Dreyfuss, 1996; Yong et al, 2002, 2004). While the controls showed that the inducible proteins functioned as endogenous proteins and were expressed at usable levels an issue became apparent when siRNAs were used to deplete target proteins in HEK293 cells. Comparative analysis of the levels of the SMN protein in HEK293 and HeLa SS6 cells transfected with siRNAs targeting SMN revealed that siRNA depletion of proteins was not as efficient in HEK293 as in HeLa SS6 cells. Furthermore, the transfection of siRNAs targeting NOP56, NOP58 and fibrillarin in HEK293 cells did not result in a change in levels or cellular localisation of the U3 snoRNA as was shown in HeLa SS6 cells (Figure 3.4). To overcome this issue numerous transfection protocols were employed including electroporation; however, these were unsuccessful. Even though the use of siRNA mediated depletion was not possible these Tet-inducible cell lines can still be used to analyse the incorporation of the FLAG-SMN or FLAG-fibrillarin into newly synthesised complexes by immunoprecipitation using anti-FLAG or anti-His antibodies. At the time of writing, numerous shRNAs are being developed by the lab to create a more efficient protein depletion system for use in HEK293 cells.

To overcome the issues of siRNA mediated depletion in HEK293 cells the inducible DD-fibrillarin reporter plasmid was developed that could be co-transfected with siRNAs into HeLa SS6 cells. Using this system it was shown that the depletion of NOP56 resulted in the accumulation of newly synthesised DD-fibrillarin in the nucleoplasm (Figure 3.19). A possible explanation for this is that DD-fibrillarin was incorporated into pre-snoRNP complexes that were unable to localise to the nucleolus, due to the lack of NOP56, resulting in the subsequent accumulation in the nucleoplasm. This is also

indicated by the fact that the depletion of NOP56 resulted in the accumulation of U3 box C/D snoRNA in the nucleoplasm (Figure 3.4).

The siRNA mediated depletions of other box C/D snoRNP factors was attempted using the DD-fibrillarin co-transfection procedure; however, the use of this procedure was limited due to the lack of DD-fibrillarin accumulation in all but the NOP56 depletion experiments. Another issue that limited the use of the DD-fibrillarin reporter plasmid was that the GFP transfection marker did not express at high enough levels to determine the level of transfection and expression of DD-fibrillarin between cells, a necessary control.

In this chapter a variety of assays, procedures and approaches were developed to analyse box C/D snoRNP biogenesis; however, not all proved viable. Despite this the use of siRNA mediated depletion of target proteins followed by FISH to detect endogenous box C/D snoRNA localisation provided a method to study potential box C/D snoRNP factors. While the use of siRNA depletion of target proteins in HEK293 cells proved problematic the inducible FLAG-SMN and FLAG-fibrillarin HEK293 cell lines can be used to analyse the incorporation of the FLAG-SMN or FLAG-fibrillarin into newly synthesised complexes by immunoprecipitation using anti-FLAG or anti-His antibodies.

Chapter four

Analysis of the localisation and interactions of fibrillar

4.1. Introduction

Fibrillarin is the methyltransferase component of the box C/D snoRNPs responsible for the 2'-*O*-methylation of rRNA residues (Tollervey et al, 1993). The box C/D snoRNPs contain three other common core proteins known as 15.5K, NOP56 and NOP58, which with fibrillarin, are stably associated onto the box C/D snoRNA. In humans, the box C/D snoRNPs are responsible for 105 specific rRNA methylation events, which are essential for the structure and function of the ribosome (Maden, 1990; Decatur and Fournier, 2002). The box C/D snoRNAs guide modification of rRNA by binding target sequences using a 1-20 nucleotide guide sequence that positions the rRNA residue to be modified at the fibrillarin active site (Kiss-Laszlo et al, 1996, 1998). Fibrillarin functions by transferring a methyl group from the co-factor AdoMet to the donor site located on the ribose sugar (Figure 4.1).

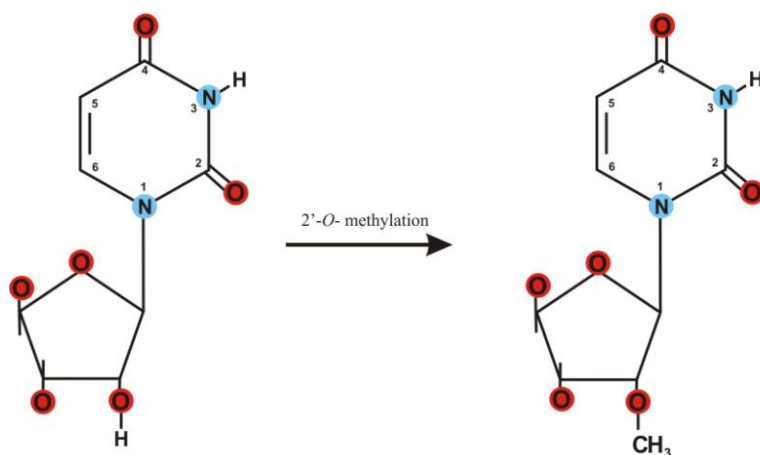


Figure 4.1: The chemical modification of rRNA by the box C/D snoRNPs

The 2'-*O*-methylation of the ribose sugar backbone of rRNA. Blue circles represent nitrogen. Oxygen is shown as red circles. Joined black lines indicate carbon. Black lines indicate covalent bonds.

Fibrillarin is a conserved protein that, in humans, contains 321 amino acids, which can be divided into the RGG motif containing domain (amino acids 1-81), the N terminal domain (amino acids 82-141) and the C terminal domain (amino acids 142-321) (Figure 4.2). The C terminal domain of fibrillarin houses the methyltransferase component that interacts with AdoMet (Wang et al, 2000; Figure 4.2 and 4.3). Interestingly, the methylated RGG motif containing domain of fibrillarin appears to be specific to eukaryotes being absent from archaea, the organism from which fibrillarin was crystallised and structure determined (Wang et al, 2000; Figure 4.3).

| | | |
|----------------------|-------|---|
| <i>H. sapiens</i> | (1) | --MKPGFSPRGGGFGGRGGFGDR--GGRGGRGGFGGG-----RGRGGGFRGRGRGGGGGGGGGGGG |
| <i>C. elegans</i> | (1) | -MGRPEFNRRGGGGGFRGGRGGDRGGSRGFGGGGRGGYGGGDRGSFGGGDRGGFRGGRGGGDRGGFRGGRGGDRGGFG |
| <i>S. cerev</i> | (1) | MSFRPGSRGGSRGGSRGFGGGRG--GSRGGARGGSRGG-----FGGRGG--RG---GARGGSRGGFG |
| | | <div> <div>β-1'</div> <div>β-2'</div> <div>β-3'</div> <div>β-4'</div> </div> |
| <i>H. sapiens</i> | (58) | GR---GGGGFHS--GGNRGRG----RGGKRGNSGKNVMEPHRHEGVFI CRGKEDALVTKNLVPGESVYGEKRVSI SE |
| <i>C. elegans</i> | (80) | GRGSPRGFGGGRG--SPRGGRGSPRGGRGGAGMRGKTVVVEPHRLGGVFIVKGKEDALATKNMVVGESVYGEKRVSVDD |
| <i>S. cerev</i> | (57) | GRGG--SRGGARG--GSRGGRG-----GAAGGARGGAKVVI EPHRHAGVYIARGKEDLVTKNMAPGESVYGEKRISVEE |
| <i>M. jannaschii</i> | (1) | -----MEDIKIKKEIFENIYEVDLGDGLKRIATKSIVKGGKKVYDEKIKIGD |
| <i>M. thermo</i> | (1) | -----MYAVKEVKGMDGVFIMNNSLLTINPN---PGVKVYGERIIDWGL |
| | | <div> <div>β-5'</div> <div>α-0</div> <div>β-1</div> <div>α-1</div> <div>β-2</div> <div>α-2</div> </div> |
| <i>H. sapiens</i> | (128) | GD-----DKIEYRAWNPFRSKLAAAILGGVDQIH IKPGA KVLVYLGAASGTTVSHVSDIVGPDGLVYAVEFSHRSGRDL |
| <i>C. elegans</i> | (159) | GAG-----SIEYRVWNPFRSKLAASIMGGLENTHIKPGTKL LYLGAASGTTVSHCSDVVGPEGIVYAVEFSHRSGRDL |
| <i>S. cerev</i> | (128) | PSKEDGVPTTKVEYRVWNPFRSKLAAGIMGGLELFIAPGKKVLVYLGAASGTSVSHVSDIVGPEGVVYAVEFSHRPGREL |
| <i>M. jannaschii</i> | (47) | E-----EYRIWNPFRSKLAAAIKGLKVXPIKRDSK LLYLGASAGTTPSHVADIADK-GIVYAI EYAPRIMREL |
| <i>M. thermo</i> | (42) | R-----EYRVWDPRFRSKLAAAILNGLRGFSLNSDSRVLYLGASAGTTASHISDITVTD-GRVYCI EFSFRMMREL |
| | | <div> <div>β-3</div> <div>α-3</div> <div>β-4</div> <div>α-4</div> <div>β-5</div> <div>α-5</div> </div> |
| <i>H. sapiens</i> | (201) | INLAKKRTNIIPVIEDARHPHKYRMLIAMVDVIFADVAQPDQTRIVALNAHTFLRNGGHFVISIKANCIDSTASAEAVFA |
| <i>C. elegans</i> | (232) | LGVAKKRPNVVPIVEDARHPHKYRMLVGMVDVIFSDVAQPDQARIVALNAQNFLRNGGHAVISIKANCIDSTAEAEAVFA |
| <i>S. cerev</i> | (208) | ISMAKKRPNIIPVIEDARHPQKYRMLIGMVDVCFADVAQPDQARIVALNSHMF LKDOGGVVISIKANCIDSTVDAETVFA |
| <i>M. jannaschii</i> | (115) | LDACAERENIIPILGDANKPQOEYANIVEKVDVIYEDVAQPNQAEILIKNAKWFLKKGXIAIKARSIDVTKDPEKIFK |
| <i>M. thermo</i> | (110) | LGVCESRKNNMAPLLEDASRPLSYLRMVEAADLVYCDVAQPDQTRIFLENMDCFLKRDGYGLIMIKARSIDVTRSPRKIFR |
| | | <div> <div>α-6</div> <div>β-6</div> <div>β-7</div> </div> |
| <i>H. sapiens</i> | (281) | SEVKKMQQENMKPQEQLTLEPYERDHA VVGVYRPPPKVKN---- |
| <i>C. elegans</i> | (312) | GEVNLKKEEFKPLEQVTLLEPYERDHA VVAVYRPFVKGKKV---- |
| <i>S. cerev</i> | (288) | REVQKLREERIKPLEQLTLEPYERDHCIVVGRYMRSGLKK---- |
| <i>M. jannaschii</i> | (195) | EQKEILEAGGFKIVDEVDI EPF EKDHVMFVGITWEGK----- |
| <i>M. thermo</i> | (190) | EEVGLKLESSGFRIIDQVGLNPEY EKDHMAVLVKRDV----- |

Figure 4.2: Protein sequence alignment of fibrillarlin from various organisms

Protein sequence alignments of fibrillarlin from *Homo sapiens* (*H. sapiens*), *Caenorhabditis elegans* (*C. elegans*), *Saccharomyces cerevisiae* (*S. cerev*), *Methanococcus jannaschii* (*M. jannaschii*) and *Methanococcus thermophiles* (*M. thermo*) (indicated to the left of the sequences). Sequences were aligned by clustalW. Identical amino acids marked by black boxes and similar amino acids marked by grey boxes. Black lines above sequence show location of β -sheets and α -helices as indicated from the archaea *M. jannaschii* fibrillarlin crystal structure (Wang et al, 2000; see Figure 4.3).

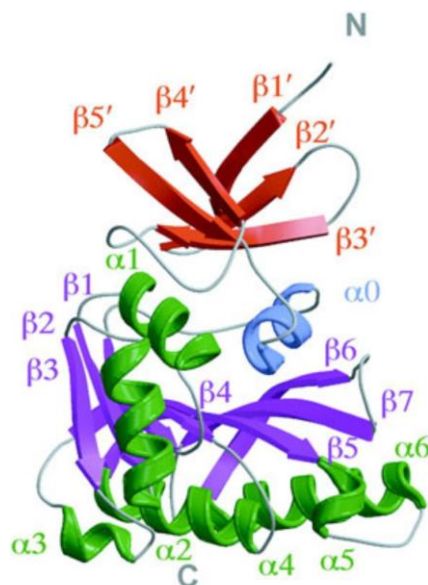


Figure 4.3: Crystal structure of archaeal fibrillarin

Ribbon diagram of fibrillarin from the archaea *Methanococcus jannaschii* with secondary structures assigned by PROCHECK and ribbon representation by MOLSCRIPT. β -sheets in the C terminal domain (β_1 – β_7) and in the N terminal domain (β_1 – β_5) are shown in magenta and red, respectively. The C terminal domain α -helices (α_1 – α_6) are shown in green while the helix (α_0) that connects the N and C terminal domains is shown in blue. Image taken from Wang et al, 2000.

Fibrillarin is found in the DFC sub-compartments of nucleolus where, as a stable component of the box C/D snoRNPs, it functions in the 2'-*O*-methylation of rRNA residues (Figure 4.1 and 4.4). The nucleolus contains three sub-compartments, the FC, DFC and GC, and is the site where synthesis and processing of pre-rRNA as well as the bulk of ribosome assembly events occur. The pre-rRNA is transcribed at the FC DFC border and, as the pre-rRNA migrates from the FC to the DFC and finally the GC, it is processed and modified. Factors that function in both early and late pre-rRNA processing, such as BMS1, are distributed throughout the nucleolus in both the DFC and GC (Wegierski et al, 2001; Figure 4.4).

Fibrillarin is also found in the Cajal bodies, which could represent box C/D snoRNPs during biogenesis (Verheggen et al, 2001; Boulon et al, 2004; Lemm et al, 2006) or box C/D scaRNPs that are localised in the Cajal bodies (Jady and Kiss, 2001)

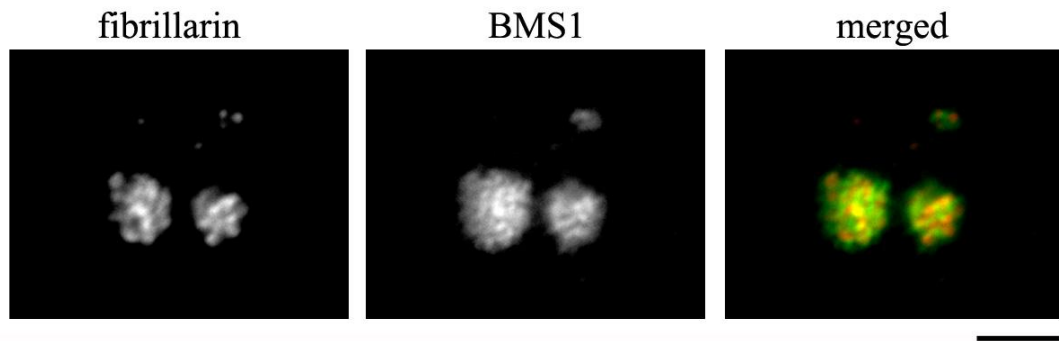


Figure 4.4: Differential nucleolar localisation of fibrillarin and BMS1

HeLa SS6 cells were grown on coverslips, transfected with plasmids that express 2 x flagged tagged fibrillarin and immunofluorescence performed using anti-FLAG antibodies. Immunofluorescence was also performed using antibodies against BMS1 to detect the nucleolus. The proteins visualised are indicated above the panel. Merged image shows fibrillarin (red) with BMS1 (green). The bar to the bottom right of the panel represents 5 μ m.

A mutational study of human fibrillarin revealed that amino acids 133-306, which closely corresponds to the C terminal domain (amino acids 142-321), were essential for the localisation to both the nucleolus and the Cajal bodies (Snaar et al, 2000). The same study also showed that the fibrillarin RGG motif containing domain (amino acids 1-80) and amino acids 81-133, which corresponds closely to the N terminal domain (amino acids 82-141), were not essential for fibrillarin localisation (Snaar et al, 2000). However, the mechanisms that result in the localisation of fibrillarin to the DFC and Cajal bodies are not clear. The region of fibrillarin spanning amino acids 133-306 could contain nucleolar and Cajal body localisation signals or alternatively be essential for association into the box C/D snoRNPs which then localise to the Cajal bodies and nucleolus.

The biogenesis of box C/D snoRNPs is a complex process that has been proposed to occur in the nucleoplasm (Terns et al, 1995; Watkins et al, 2004; Boulon et al, 2004), Cajal bodies (Verheggen et al, 2001; Boulon et al, 2004; Lemm et al, 2006) and cytoplasm (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007). Box C/D snoRNP biogenesis proceeds in a large, dynamic complex known as the pre-snoRNP that consists of the snoRNA, common core proteins and assembly, transport and RNA processing factors (Boulon et al, 2004; Watkins et al, 2004, 2007; McKeegan et al,

2007). The assembly factors TIP48, TIP49, BCD1, TAF9, NOP17 and NUFIP are vital for box C/D snoRNP biogenesis and have been shown to extensively interact with the common core proteins (McKeegan et al, 2007). It has been proposed that the box C/D snoRNP assembly factors function in the early recruitment of the common core proteins to the pre-snoRNP (Watkins et al, 2004, 2007; McKeegan et al, 2007; Boulon et al, 2008). Fibrillarin interacts with all of the above assembly factors; however, the specific function of the assembly factors is not known.

The importance of the interactions between fibrillarin and the box C/D snoRNP assembly factors for association with the box C/D snoRNP has not been analysed. Furthermore, the contribution of the various domains of fibrillarin for incorporation into the box C/D snoRNP has not been fully characterised. To gain a greater understanding of the association of fibrillarin with the box C/D snoRNPs a series of fibrillarin deletion mutants were produced and a variety of experimental approaches utilised. These included assays to determine which domains of fibrillarin were required for localisation, association with snoRNPs and interactions with the box C/D snoRNP assembly factors. By further analysing the roles of the various domains of fibrillarin the association with the box C/D snoRNP could be further defined.

4.2. Results

4.2.1. Construction of human fibrillarin deletion mutants

To further characterise the association of fibrillarin with the box C/D snoRNP a series of human fibrillarin deletion mutants were constructed. These deletion mutants were used in a variety of experiments to further define the roles and interactions of the individual domains of fibrillarin.

The fibrillarin deletion mutants were constructed based, in part, upon analysis of the crystal structure of the archaea, *Methanococcus jannaschii* (*M. jannaschii*), fibrillarin (Wang et al, 2000; see Figure 4.3). Analysis of the crystal structure revealed two main globular domains, the N terminal and C terminal domains, which are separated by alpha helix 0. The N terminal domain consists of five β -sheets while the larger C terminal domain has seven β -sheets and seven α -helices (Wang et al, 2000; see Figure 4.3). Sequence alignments were performed between human and *M. jannaschii* fibrillarin, which showed that there was a high degree of identity between these two proteins (Figure 4.2). The locations of the fibrillarin secondary structures, as shown by the *M. jannaschii* fibrillarin crystal structure, were then assigned to the human fibrillarin sequence (Figure 4.2). One major difference between archaea and eukaryotic fibrillarin was the presence of an RGG motif containing domain located upstream of the N terminal domain in the eukaryotic protein.

Based upon the identification of the locations of the N and C terminal domains and the additional eukaryotic specific RGG containing domain five fibrillarin deletion mutants were constructed by PCR amplification of human fibrillarin cDNA. These included an N and C terminal domain mutant that lacked the RGG domain (N-C-fibrillarin; amino acids 82-321), an RGG domain and N terminal domain mutant that lacked the C terminal domain (R-N-fibrillarin; amino acids 1-141), a RGG domain mutant that lacked both the N and C terminal domains (R-fibrillarin; amino acids 1-81) and a N terminal domain mutant that lacked the RGG domain and C terminal domain (N-fibrillarin; amino acids 82-141) (see Figure 4.5). In addition a full length fibrillarin was produced to be used as a positive control. The amplified cDNA of fibrillarin and the

deletion mutants was cloned into plasmids that inserted N terminal GFP tags (for mammalian expression) and pET plasmids for expression in *E. coli* and to produce *in vitro*-translated [³⁵S] methionine-labelled proteins from reticulocyte lysate.

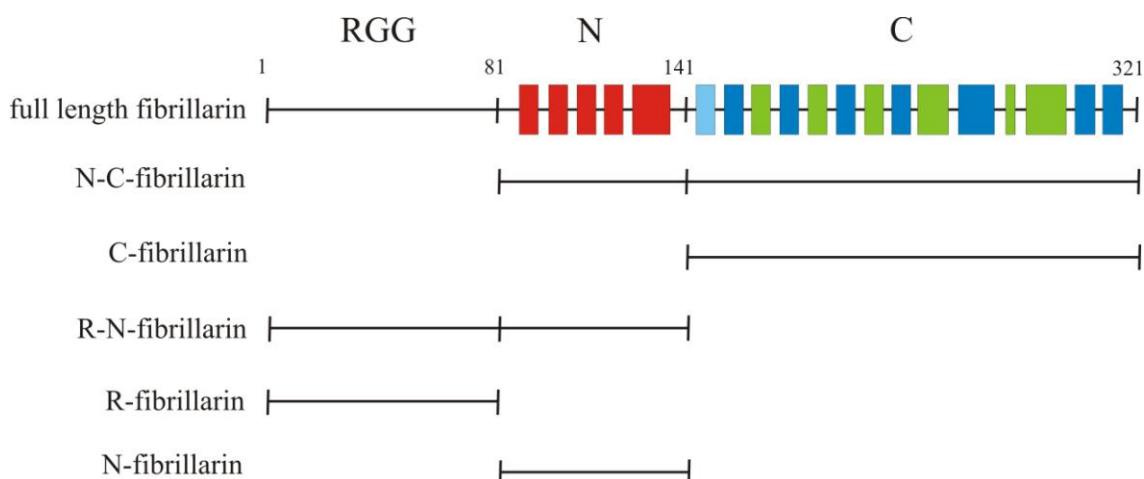


Figure 4.5: Fibrillarlin deletion mutants

Schematic representation of human fibrillarlin deletion mutants. The full length fibrillarlin is composed of a RGG motif containing domain (RGG; amino acids 1-81), N terminal domain (N; amino acids 82-141) and C terminal domain (C; amino acids 142-321), which are indicated above the panel. Secondary structures are indicated by coloured blocks, which correspond to Figure 4.3. β-sheets in the C terminal domain (β₁–β₇) and in the N terminal domain (β₁–β₅) are shown in magenta and red, respectively. The C terminal domain α-helices (α₁–α₆) are shown in green while the α-helix (α₀), which connects the N and C terminal domains, is shown in blue. Numerical values above panel correspond to amino acid number. The five fibrillarlin deletion mutants are shown below the full length fibrillarlin. Name of deletion mutant is shown to the left of the panel.

4.2.2. Analysis of the localisation and association, with the box C/D snoRNPs, of the fibrillarlin deletion mutants

To determine which domains of fibrillarlin are required for correct cellular localisation HeLa SS6 cells were transfected with plasmids that expressed full length fibrillarlin and the deletion mutants with N terminal GFP fusion tags. Twenty-four hours after transfection the cells were fixed and the localisation of the GFP tagged fibrillarlin and deletion mutants was viewed by widefield microscopy.

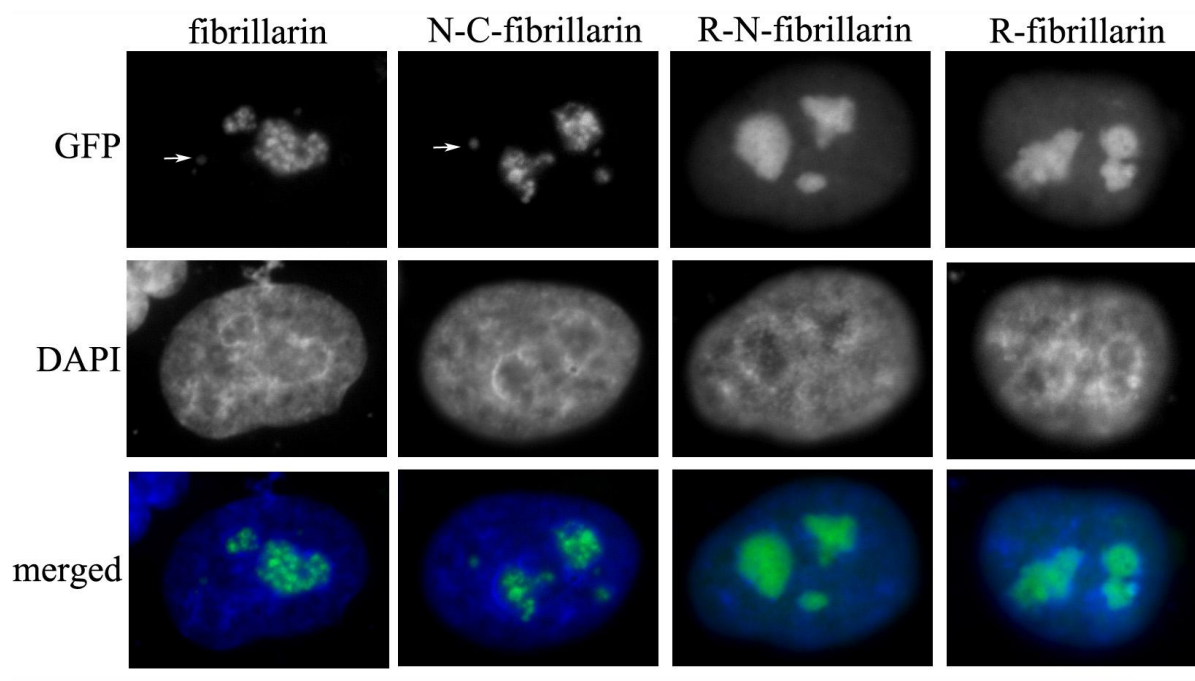


Figure 4.6: The localisation of GFP-fibrillarin deletion mutants

HeLa SS6 cells were transfected with plasmids encoding GFP fusions of full length fibrillarin and the deletion mutants. After twenty-four hours cells were fixed and stained with DAPI to identify the nucleoplasm. The GFP-fibrillarin deletion mutant plasmid transfected is shown above each lane (see Figure 4.5). Top row shows the GFP-fibrillarin deletion mutant proteins, the middle row shows the DAPI image and the bottom row is a merged image of GFP-fibrillarin mutant proteins (green) with the DAPI image (blue). The white arrows indicate a Cajal body. The bar to the bottom right of the panel represents 5 μ m.

The full length GFP-fibrillarin was found predominately in discrete sub-compartments of the nucleolus, indicating localisation to the DFC (Figure 4.6 (Figure 4.4)). All of the fibrillarin deletion mutants tested also localised to the nucleolus; however, only the GFP-N-C-fibrillarin deletion mutant was found specifically in the DFC (Figure 4.6). The GFP-R-N-fibrillarin and GFP-R-fibrillarin deletion mutants were distributed throughout the nucleolus, which suggests localisation to both the DFC and GC.

The full length GFP-fibrillarin was also found enriched in Cajal bodies (Figure 4.5), which is characteristic of wild type fibrillarin. Of the deletion mutants only the GFP-N-C-fibrillarin was found in Cajal bodies. While the GFP-R-N-fibrillarin and GFP-R-fibrillarin deletion mutants were not seen in Cajal bodies they were distributed throughout the nucleoplasm, which was not seen with full length GFP-fibrillarin. The localisation of GFP-R-N-fibrillarin and GFP-R-fibrillarin deletion mutants to the nucleoplasm could, however, be due to the over expression of these proteins resulting in non-specific accumulation.

Unfortunately, it was not possible to determine the localisation of the N-fibrillarin deletion mutant due to the inability to successfully clone the cDNA into the GFP expression plasmid. It was also not possible to determine the localisation of the GFP-C-fibrillarin deletion mutant; however, this was due to the expression level being too low to be detected by either widefield microscopy or even Western blot analysis using anti-GFP or anti-fibrillarin antibodies (data not shown).

The fibrillarin deletion mutants were also created to establish which domains of fibrillarin are essential for association with the box C/D snoRNPs. To determine this the GFP tagged fibrillarin deletion mutants were expressed in HeLa SS6 cells, whole cell extracts produced, immunoprecipitation assays performed using anti-GFP antibodies and co-precipitating factors, such as box C/D snoRNA, analysed. These experiments were, however, unsuccessful and no co-precipitation of box C/D snoRNA was witnessed even with the full length GFP-fibrillarin control. Two possible explanations are that the anti-GFP antibodies were inefficient at immunoprecipitating the GFP tagged full length fibrillarin and deletion mutants and / or the expression level of the GFP tagged full length fibrillarin and deletion mutants was too low to result in the precipitation of enough material to be detected.

4.2.3. Fibrillarin interacts directly with the box C/D snoRNP assembly factors

Previous studies have shown that fibrillarin interacts with the box C/D snoRNP assembly factors TIP48, TIP49, BCD1, NOP17, TAF9 and NUFIP (McKeegan et al, 2007). However, one issue with the *in vitro* interaction assays used to determine these interactions was the use of *in vitro*-translated ³⁵S-labelled fibrillarin. As the ³⁵S-labelled fibrillarin was produced from reticulocyte lysate it is possible that additional factors, that could bridge interactions, may be present in the assays making it difficult to determine if the interactions between fibrillarin and the box C/D snoRNP assembly factors are direct.

To determine whether fibrillarin interacts directly with the box C/D snoRNP assembly factors *in vitro* interaction assays were performed using recombinant human fibrillarin, expressed and purified from *E. coli*. Unfortunately, due to solubility issues, it was not possible to express and purify full length fibrillarin fused to an N terminal 6 x His tag (His) efficiently. It was possible, however, to express and purify the N-C-fibrillarin deletion mutant fused with an N terminal His tag and this protein was used in the *in vitro* interaction assays (His-N-C-fibrillarin). The N terminal His tag allowed N-C-fibrillarin to be specifically identified in a Western blot assay using anti-His antibodies.

The box C/D snoRNP assembly factors TIP48, TIP49, NOP17 and NUFIP were all expressed and purified with N terminal GST fusion tags from *E. coli* (GST-TIP48, GST-TIP49, GST-NOP17 and GST-NUFIP; McKeegan et al, 2007; Figure 4.7 A). Unfortunately, it was not possible to express and purify full length BCD1 and therefore a BCD1 deletion mutant, consisting of amino acids 1-360 was expressed and purified fused to an N terminal GST tag. As the GST-BCD1 deletion mutant was previously used to identify the interaction with fibrillarin (McKeegan et al, 2007) there was no issue with using this protein to analyse the interaction with fibrillarin.

The GST tagged assembly factors or GST alone (negative control) were immobilised on glutathione Sepharose beads and incubated with His-N-C-fibrillarin. After two hours incubation at 4°C the glutathione Sepharose beads were washed, bound proteins

separated on a 12 % SDS-PAGE gel and Western blot assays performed using anti-His antibodies (Figure 4.7 B).

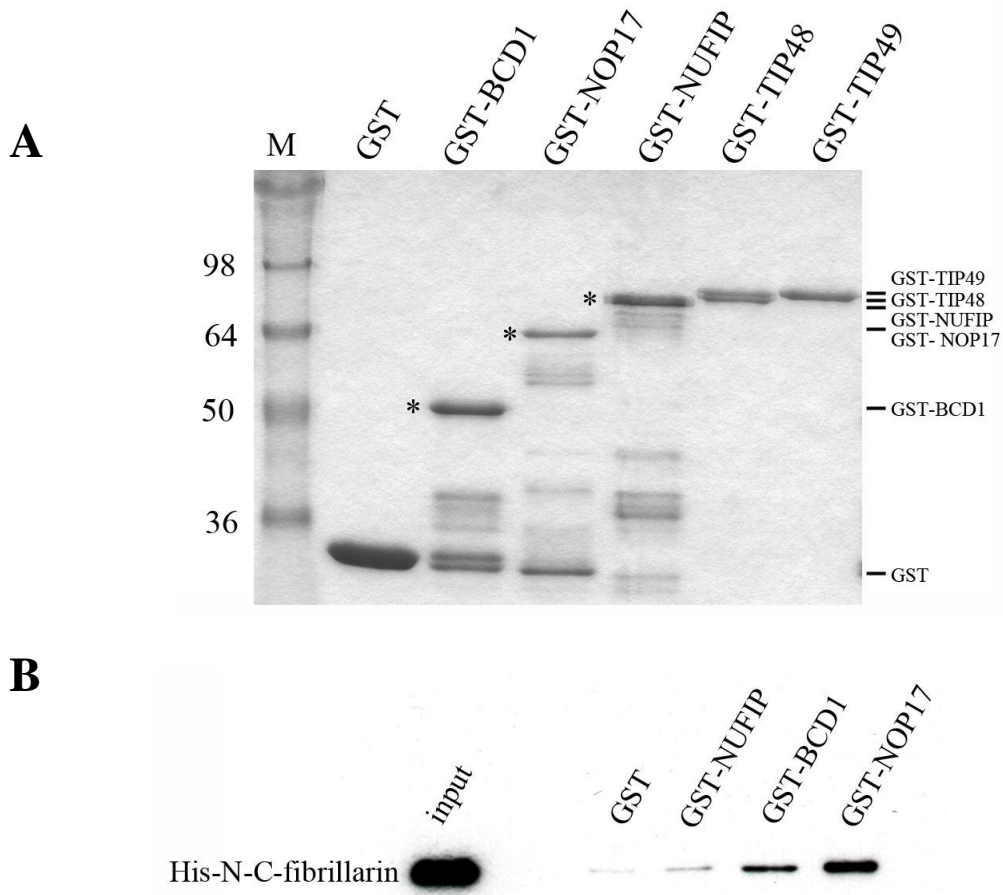


Figure 4.7: Interaction assay between fibrillarin and the box C/D snoRNP assembly factors NUFIP, BCD1 and NOP17

(A) Recombinant GST, GST-BCD1 (deletion mutant amino acids 1-360), GST-NOP17, GST-NUFIP, GST-TIP48 and GST-TIP49 proteins were separated on a 12 % SDS-PAGE gel and visualised by Coomassie blue staining. The identity of the protein loaded is indicated at the top of the panel. The migration of individual proteins is indicated to the right of the panel. M: molecular marker. The size of the molecular marker is shown to the left of the panel in kilodaltons. The asterisk indicates the full length protein where appropriate

(B) Interaction of fibrillarin with box C/D snoRNP assembly factors. GST, GST-NUFIP, GST-BCD1 and GST-NOP17 were immobilised on glutathione Sepharose beads and incubated with His-N-C-fibrillarin. Bound proteins were separated on a 12 % SDS-PAGE gel and His-N-C-fibrillarin detected by Western blot analysis using anti-His antibodies. Input represents 10 % of His-N-C-fibrillarin used in interaction assay. The immobilised GST proteins used are shown above the panel.

Immobilised GST-BCD1 and GST-NOP17 retained His-N-C-fibrillarin at levels above GST alone (Figure 4.7). His-N-C-fibrillarin was also retained by immobilised GST-NUFIP above the levels seen with GST alone; however the retention was low. The weak retention of His-N-C-fibrillarin by GST-NUFIP was repeatedly seen, which suggests a weak interaction between these proteins. These experiments indicate that fibrillarin interacts directly with the box C/D snoRNP assembly factors BCD1, NOP17 and NUFIP.

In vitro interaction assays were also performed to determine whether the interaction of fibrillarin with the box C/D snoRNP assembly factors TIP48 and TIP49 were direct. The assays were performed as described above, however, unlike the previous experiment the interactions were visualised by direct Coomassie blue staining of the SDS-PAGE gel. Assay shown in Figure 4.8 performed by K.S.McKeegan, Newcastle University, UK.

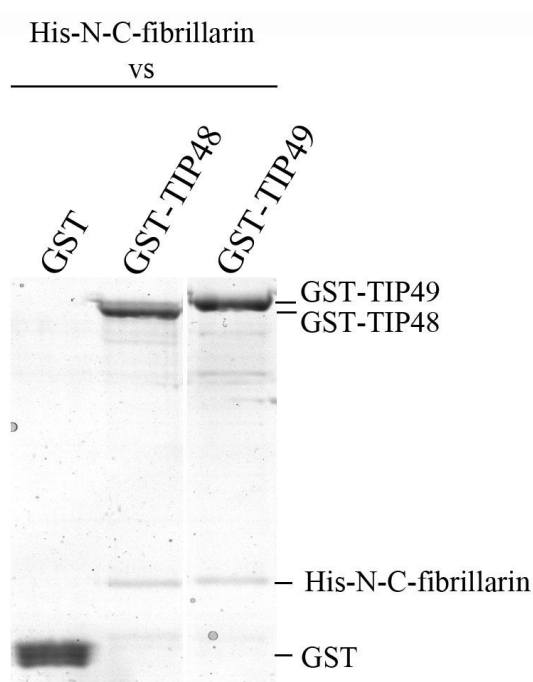


Figure 4.8: Interaction assay between fibrillarin and the box C/D snoRNP assembly factors TIP48 and TIP49

GST, GST-TIP48 and GST-TIP49 were immobilised on glutathione Sepharose beads and incubated with His-N-C-fibrillarin. Bound proteins were separated on a 12 % SDS-PAGE gel and interactions visualised by Coomassie blue staining. The immobilised GST proteins are shown above the panel. The migration of individual proteins used is indicated to the right of panel. Result kindly provided by K.S. McKeegan, Newcastle University, UK.

Immobilised GST-TIP48 and GST-TIP49 both retained His-N-C fibrillarin above the levels seen with GST alone (Figure 4.8). This indicates that fibrillarin interacts directly with both TIP48 and TIP49.

4.2.4. The box C/D snoRNP assembly factors interact with the C terminal domain of fibrillarin

To determine which domains of fibrillarin the box C/D snoRNP assembly factors NUFIP, BCD1, NOP17, TIP48 and TIP49 interact with *in vitro* protein interaction assays were performed using the fibrillarin deletion mutants. Unfortunately, it was not possible to express and purify all the fibrillarin deletion mutants as recombinant proteins due to solubility issues. Therefore, full length fibrillarin and the deletion mutants were expressed using reticulocyte lysate in the presence of [³⁵S] methionine. The GST tagged assembly factors were immobilised on glutathione Sepharose beads and incubated with the *in vitro*-translated ³⁵S-labelled full length fibrillarin and deletion mutant proteins. Bound proteins were separated on a 12 % SDS-PAGE gel and results visualised by autoradiography.

Immobilised GST-NUFIP retained *in vitro*-translated ³⁵S-labelled full length fibrillarin and all the deletion mutant proteins above the levels seen with GST alone (Figure 4.9). GST-NUFIP retained the full length fibrillarin and the N-C-fibrillarin and C-fibrillarin deletion mutants at a higher level than the R-N-fibrillarin and N-fibrillarin, which indicates that NUFIP interacts most strongly with the C terminal domain of fibrillarin. As GST-NUFIP retains R-N-fibrillarin and N-fibrillarin above the levels of GST alone, albeit at a low intensity, this suggests that NUFIP may interact weakly with N terminal domain of fibrillarin.

Immobilised GST-BCD1 retained *in vitro*-translated ³⁵S-labelled full length fibrillarin and the N-C-fibrillarin and C-fibrillarin deletion mutants above the levels seen with GST alone (Figure 4.9). As GST-BCD1 only retained fibrillarin deletion mutants that contained the C terminal domain this indicates that BCD1 interacts with this domain. As the retention of the N-C-fibrillarin and C-fibrillarin deletion mutants by GST-BCD1

was lower than with full length fibrillarin this suggests that the whole fibrillarin protein is required for an efficient interaction.

Immobilised GST-NOP17 retained *in vitro*-translated ^{35}S -labelled full length fibrillarin and the N-C-fibrillarin and C-fibrillarin deletion mutants above the levels of GST alone (Figure 4.9). As GST-NOP17 did not retain fibrillarin deletion mutants that lacked the C terminal domain this indicates that NOP17 interacts specifically with the C terminal domain of fibrillarin.

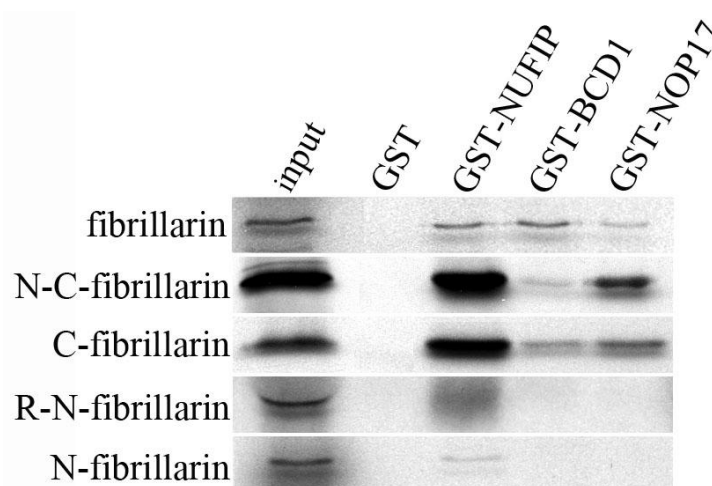


Figure 4.9: Interaction assay between the fibrillarin deletion mutants and BCD1, NOP17 and NUFIP

GST, GST-NUFIP, GST-BCD1 and GST-NOP17 were immobilised to glutathione Sepharose beads and incubated with *in vitro*-translated ^{35}S -labelled full length fibrillarin and deletion mutant proteins. Bound proteins were separated on a 12 % SDS-PAGE gel and visualised by autoradiography. Input represents 10 % of *in vitro*-translated ^{35}S -labelled protein used in interaction assay. The identity of GST tagged proteins used is indicated above each lane. The identity of the radiolabelled proteins used is indicated to the left of the panel. A single exposure was used for each *in vitro*-translated ^{35}S -labelled protein.

Immobilised GST-TIP48 and GST-TIP49 retained the *in vitro*-translated ^{35}S -labelled N-C-fibrillarin and C-fibrillarin deletion mutant proteins above levels seen with GST alone (Figure 4.10). There is a low level retention of the N-R-fibrillarin deletion mutant protein by GST-TIP48 and GST-TIP49 and no retention of the N-fibrillarin. As both immobilised GST-TIP48 and GST-TIP49 only efficiently retained the *in vitro*-translated ^{35}S -labelled fibrillarin deletion mutant proteins containing the C domain this indicates that TIP48 and TIP49 interact with the C terminal domain of fibrillarin.

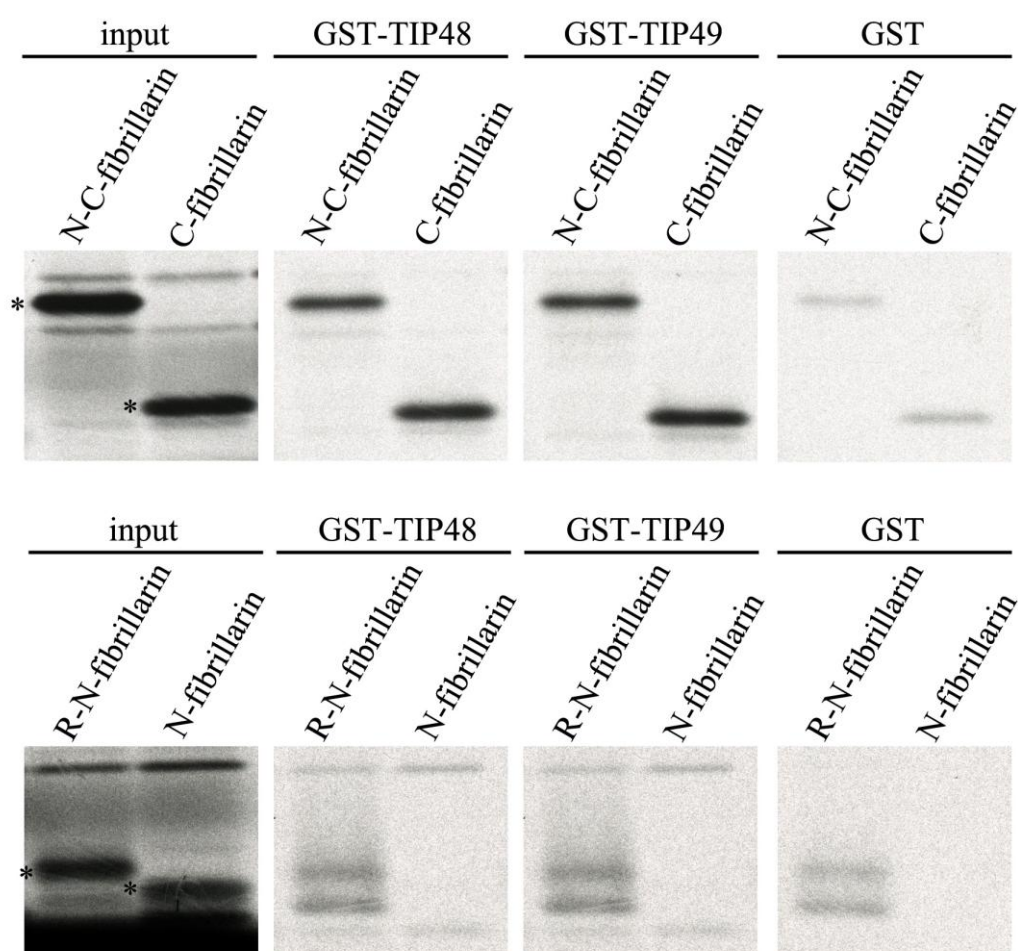


Figure 4.10: Interaction assay between the fibrillarin deletion mutants and the box C/D snoRNP assembly factors TIP48 and TIP49

GST, GST-TIP48 and GST-TIP49 were immobilised onto glutathione Sepharose beads and incubated with the *in vitro*-translated ^{35}S -labelled full length fibrillarin and deletion mutant proteins. Bound proteins were separated on a 12 % SDS-PAGE gel and visualised by autoradiography. Input represents 10 % of *in vitro*-translated ^{35}S -labelled protein used in interaction assay. The identity of GST tagged proteins and radiolabelled proteins used are indicated above each lane. A single exposure is used for each *in vitro*-translated ^{35}S -labelled protein. The asterisk indicates the full length proteins.

4.3. Discussion

In this chapter a series of human fibrillarin deletion mutants were constructed to further analyse fibrillarin and the interactions necessary for its association with the box C/D snoRNPs. The fibrillarin deletion mutants were constructed based upon analysis of the crystal structure of *M. jannaschii* fibrillarin (Wang et al, 2000), which contained two main globular domains, the N and C terminal domains. Protein sequence alignments were performed between human and *M. jannaschii* fibrillarin and the secondary structures assigned to human fibrillarin. As eukaryotic fibrillarin contains an additional domain containing an RGG motif upstream of the N terminal domain deletion mutants were also created based on the location of this domain.

To determine which domains of fibrillarin were required for correct cellular localisation the fibrillarin deletion mutants were expressed with N terminal GFP tags in HeLa SS6 cells and their localisation viewed by widefield microscopy (Figure 4.6). The GFP-N-C-fibrillarin deletion mutant localised to the DFC of the nucleolus and to the Cajal bodies. This is the typical localisation of wild type fibrillarin and indicates that the missing RGG containing domain (amino acids 1-81) of GFP-N-C-fibrillarin is not essential for localisation. Interestingly, the GFP-R-fibrillarin, which only consists of the RGG containing domain, localised throughout the nucleolus and the nucleoplasm. The GFP-N-C-fibrillarin and GFP-R-fibrillarin data is in agreement with previous studies which show that while the fibrillarin RGG motif containing domain localises to the nucleolus on its own it is not required for fibrillarin localisation (Snaar et al, 2000).

The GFP-R-N-fibrillarin deletion mutant, which lacked the C terminal domain, did not localise to Cajal bodies and was found throughout the nucleoplasm and nucleolus (Figure 4.6). As fibrillarin localisation has previously been shown not to require the RGG motif containing domain this indicates that the C terminal domain is required for localisation. This is in agreement with previous studies that have shown that a region of fibrillarin spanning amino acids 133-306, which corresponds closely with the C terminal domain used in this study (amino acids 142-321), is essential for localisation (Snaar et al, 2000). Unfortunately, it was not possible to obtain localisation data for the C-fibrillarin and N-fibrillarin deletion mutants, which prevented further analysis of the

contribution of the N and C terminal domains to cellular localisation. Whether the C terminal domain of fibrillarin contains localisation signals or is important for association of fibrillarin with the box C/D snoRNPs is not clear using this approach alone and further study is required.

Another approach used to analyse fibrillarin association with the box C/D snoRNPs was to investigate which domains of fibrillarin are required to interact with the box C/D snoRNP assembly factors. Fibrillarin has previously been shown to interact with the box C/D snoRNP biogenesis factors, BCD1, NOP17, TAF9, NUFIP, TIP48 and TIP49 (McKeegan et al 2007); however, this was determined using *in vitro*-translated ³⁵S-labelled fibrillarin. Before determining which domains of fibrillarin these assembly factors interact with it was first necessary to confirm that these interactions were direct. *In vitro* interaction assays were performed between recombinant His-N-C-fibrillarin and GST-BCD1, GST-NOP17 GST-NUFIP, GST-TIP48 and GST-TIP49. It was not possible to test whether the interaction between fibrillarin and TAF9 was direct as GST-TAF9 could not be expressed and purified at a satisfactory level. These *in vitro* interaction assays showed that fibrillarin interacts directly with BCD1, NOP17, NUFIP, TIP48 and TIP49 (Figure 4.7 and 4.8). The fibrillarin deletion mutants were then used in a series of *in vitro* interaction assays to map the box C/D snoRNP assembly factor binding domains. As it was not possible to express and purify the fibrillarin deletion mutants as recombinant proteins they were expressed as *in vitro*-translated ³⁵S-labelled proteins. These assays indicate that the box C/D snoRNP assembly factors interact with the C terminal domain of fibrillarin. There were a couple of exceptions, for instance in addition to the C terminal domain NUFIP also interacts, albeit weakly, with the N terminal domain of fibrillarin. Furthermore, while BCD1 only interacts with fibrillarin deletion mutants containing the C terminal domain the interaction was not as strong as with full length fibrillarin.

The crystal structure of the interaction of the archaeal homologue of NOP56 and NOP58 (Nop5) with fibrillarin revealed that this interaction is mediated by α -helices 2-3 and β -sheet 3 of fibrillarin, which are all located in the fibrillarin C terminal domain (Aittaleb et al 2003). As human fibrillarin has been shown to interact with NOP56 (McKeegan et al, 2007) *in vitro* interaction assays were performed using the *in vitro*-translated ³⁵S-labelled fibrillarin deletion mutants and recombinant NOP56. These

assays, however, did not produce consistent results. Based on the Nop5 interaction data (Aittaleb et al 2003) it is possible that the C terminal domain of fibrillarin mediates the interaction between fibrillarin and NOP56.

To determine which domains of fibrillarin are essential for the association with the box C/D snoRNPs the GFP tagged fibrillarin deletion mutants were expressed in HeLa SS6 cells, cell extracts produced, and anti-GFP antibodies used in immunoprecipitation assays. Co-precipitating factors such as box C/D snoRNA were then analysed. These experiments were, however, unsuccessful due either to poor expression of the GFP tagged full length fibrillarin and deletion mutants and / or the anti-GFP antibodies used were not suitable for immunoprecipitation assays. In order to overcome these issues Tet inducible fibrillarin deletion mutant HEK293 cell lines were designed and are currently being developed by the lab. A HEK293 cell line that expresses full length fibrillarin with a N terminal 2 x FLAG 6 x His tag (FLAG-fibrillarin) has already been developed and successfully used in immunoprecipitation and immunofluorescence assays in this study (Chapter 3.2.6). The use of inducible 2 x FLAG 6 x His tagged fibrillarin deletion mutant HEK293 cell lines should overcome the issues experienced with immunoprecipitation assays using the GFP-fibrillarin deletion mutants.

Overall the data presented indicates that the C terminal domain of fibrillarin is essential for both fibrillarin localisation and for interactions with the box C/D snoRNP assembly factors. It is possible that the interaction of the C terminal domain of fibrillarin with the assembly factors mediates the association with the box C/D snoRNPs, which then localise to the Cajal bodies and the nucleolus. Further experiments are, however, required to confirm that the C terminal domain of fibrillarin is essential for the association of fibrillarin with the box C/D snoRNPs.

Chapter five

The role of the SMN complex in box C/D snoRNP biogenesis

5.1. Introduction

The multiprotein SMN complex functions in the assembly of the snRNPs (Yong et al, 2002, 2004) and has also been implicated in the biogenesis of telomerase RNPs (Bachand et al, 2002), miRNPs (Mourelatos et al, 2002) and snoRNPs (Jones et al, 2001; Pellizzoni et al, 2001). The SMN complex has been under intense investigation as mutation of the gene encoding the SMN protein is linked to the neurodegenerative disease SMA.

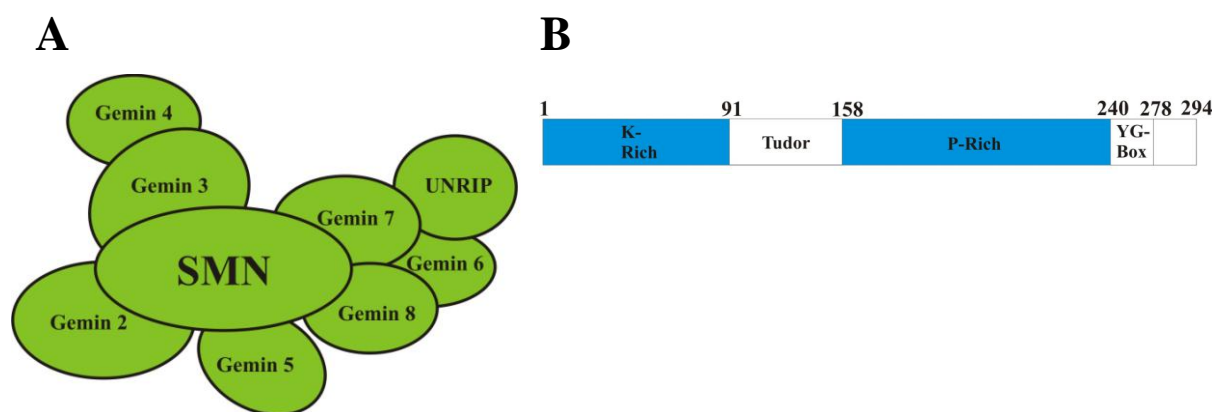


Figure 5.1: The SMN complex

(A) Schematic diagram of the SMN complex. SMN complex components are shown as green ovals. It is possible that multiple SMN proteins self oligomerise to form larger complexes.

(B) Schematic diagram of the domain organisation of the SMN protein. SMN contains an N terminal domain K-rich (lysine) sequence, a central Tudor domain and C terminal domain P-rich (proline) sequences and YG-box (tyrosine glycine). Numerical values above the panel indicate the amino acid numbers.

The SMN complex is composed of SMN, Gemin2-8 and the UNRIP factor. The SMN protein is the central component of the SMN complex and interacts directly with Gemin2, 3, 5, 7 and 8 (Figure 5.1 A). Not all the SMN complex proteins, however, interact with the SMN protein, for instance Gemin4 and 6 associate with the SMN complex through interactions with Gemin3 and 7. UNRIP associates with the complex by interacting with Gemin6 and 7 and is exclusively found in cytoplasmic SMN complexes (Carissimi et al, 2005). Gemin8 also interacts with Gemin6 and 7. The exact stoichiometry of the SMN complex is not clear and it has been proposed that the SMN protein self oligomerises to form a much larger complex (Paushkin et al, 2002). The

SMN complex is found throughout the cytoplasm and in the nucleus enriched in Cajal bodies and Gems (Liu and Dreyfuss, 1996).

The human SMN protein is of 294 amino acids in size and contains a number of conserved motifs (Figure 5.1 B). These include an N terminal domain RNA binding K-rich sequence and a central Tudor domain that mediates interactions with RGG motifs of proteins including coilin and Sm proteins (Selenko et al, 2001; Whitehead et al, 2002). In addition the SMN protein contains an YG-box, which has been implicated in self oligomerisation (Lorson et al, 1998). The Gemin and UNRIP proteins also contain evolutionary conserved motifs, for instance Gemin3 contains a DEAD box RNA helicase motif (Charroux et al, 1999) while Gemin5 contains 13 WD repeats in the N terminal domain and coiled-coil motifs in the C terminal domain (Gubitz et al, 2002). Gemin7 contains several RGG motifs (Baccon et al, 2002) and UNRIP contains a number of GH-WD motifs (Carissimi et al, 2005).

The most characterised role of the SMN complex is in the biogenesis of the snRNPs, which function in pre-mRNA splicing as part of the spliceosome. The SMN complex has a number of proposed roles and the best defined of these is the association of the Sm core proteins (B/B', D1, D2, D3, E, F and G) onto pre-snRNA to form a stable heptameric ring known as the Sm core (Figure 5.2; Charroux et al, 1999; Friesen et al, 2001; Brahms et al, 2001; Gubitz et al, 2002). In the cytoplasm the SMN complex functions in Sm core formation by binding preformed Sm protein complexes (D3/B, D1/D2 and E/F/G), which are associated with pICln (a component of PRMT5 complex), and then catalysing the formation of the Sm core on pre-snRNA (Chari et al, 2008).

The SMN complex has also been implicated in the nuclear import of pre-snRNPs. The nuclear import of the pre-snRNPs occurs through the utilisation of one of two NLS that are formed in the cytoplasm, the m₃G cap and Sm core. The SMN complex has known roles in the formation of the Sm core but has also been implicated in the recruitment of the hypermethylase TGS1 that converts the snRNA 5' m⁷G cap into an m₃G cap (Mouaikel et al, 2003). In the m₃G dependent import pathway Snurportin1 interacts with the m₃G cap of the pre-snRNA and the import factor importin β, which results in the nuclear import of the pre-snRNP (Huber et al, 1998). SMN has been proposed to function in the Sm core dependent import pathway as SMN has also been shown to

interact with importin β and can mediate nuclear import in the absence of Snurportin1 (Narayanan et al, 2002).

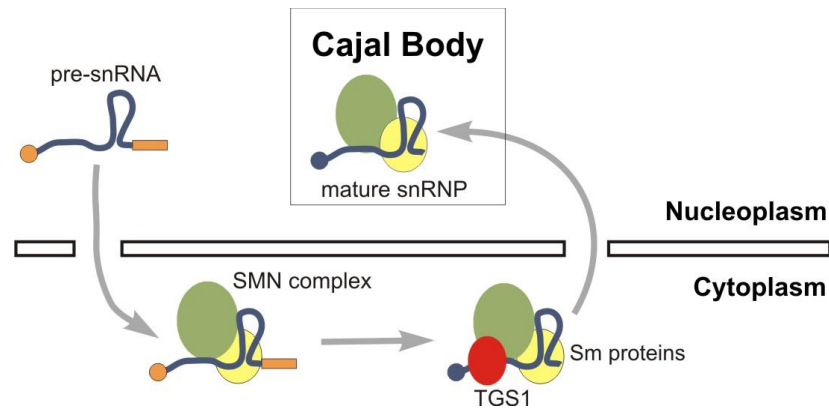


Figure 5.2: The role of the SMN complex in snRNP biogenesis

A dark grey line represents the snRNA with the 5' m^7G cap shown as an orange circle and the 3' extended sequence as an orange rectangle. The Sm core proteins are represented by a yellow oval. The SMN complex is represented by a green oval. TGS1 is shown as a red oval. The 5' m_3G cap of mature snRNA is shown as by a dark grey circle. The Cajal body is represented as a grey outlined box. The nuclear membrane is shown by black outlined rectangles. Grey arrows indicate the progression of biogenesis. Figure adapted from image provided by N.J.Watkins, Newcastle University, UK.

The SMN complex has also been linked to the biogenesis of the box C/D snoRNPs, which predominately function in the chemical modification of rRNA (Kiss-Laszlo et al, 1996). However, a few box C/D snoRNPs, such as U3 and U8, assist in the processing of pre-rRNA (Sharma and Tollervey, 1999; Kiss et al, 2002). The box C/D snoRNPs are stable RNA protein complexes which consist of four evolutionary conserved common core proteins known as 15.5K, NOP56, NOP58 and fibrillarin. The fibrillarin core protein is the methyltransferase component of the complex, which is responsible for the 2'-*O*-methylation of target rRNA residues (Tollervey et al, 1993).

The box C/D snoRNPs are found in the nucleolus where the synthesis, processing and modification of pre-rRNAs occur, as well as the bulk of ribosome assembly events. The nucleolus contains three sub-compartments, the FC, DFC and GC. The box C/D snoRNPs that function in the chemical modification of rRNA and early pre-rRNA processing, such as the U8 snoRNP (Peculis and Steitz, 1993), are found in the discrete DFC sub-compartments of the nucleolus (Figure 5.3). On the contrary, factors that

function in both early and late pre-rRNA processing such as the U3 box C/D snoRNP are found distributed throughout the nucleolus in both the DFC and GC (Filipowicz et al, 1999; Lafontaine and Tollervey, 2001; Granneman, et al, 2004). Box C/D snoRNA is also found in the Cajal bodies and this, more than likely, represents pre-snoRNPs during biogenesis (Verheggen et al, 2001; Boulon et al, 2004; Lemm et al, 2006).

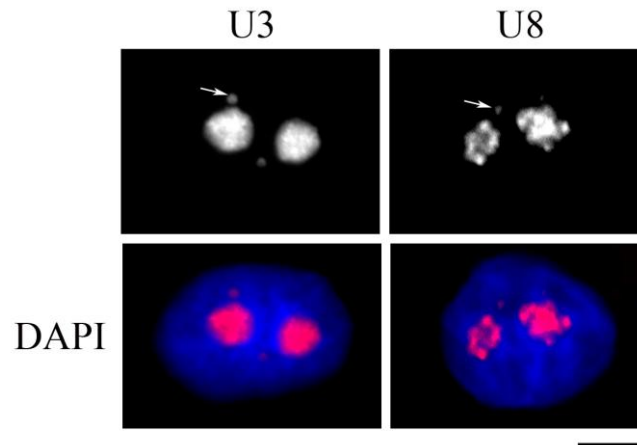


Figure 5.3: Differential nucleolar localisation of the box C/D snoRNPs

HeLa SS6 cells were grown on coverslips, fixed and FISH performed using fluorescent probes specific for the U3 and U8 box C/D snoRNAs. DAPI staining of DNA was used to indicate the nucleoplasm. The top row shows the U3 or U8 snoRNA and the bottom row a merged image of U3 or U8 (red) with the DAPI staining (blue). The white arrow indicates a Cajal body. The bar to the bottom right of the panel represents 5 μ m.

The box C/D snoRNAs are transcribed in the nucleoplasm by RNA polymerase II and form a large, multiprotein pre-snoRNP complexes that contain snoRNA, common core proteins and assembly, RNA processing and transport factors (Watkins et al, 2004, 2007; McKeegan et al, 2007; Boulon et al, 2004, 2008). Numerous interactions between the box C/D snoRNP assembly factors BCD1, NOP17, NUFIP, TAF9, TIP48 and TIP49 and the common core proteins have been identified and it has been proposed that these interactions are required for the recruitment and association of the common core proteins with the box C/D snoRNA (Watkins et al, 2004, 2007; McKeegan et al, 2007). The site(s) of box C/D snoRNP biogenesis is not clear with the nucleoplasm (Watkins et al, 2004; 2007; McKeegan et al, 2007), Cajal bodies (Boulon et al, 2007; Verheggen et al, 2002) and cytoplasm (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007) all implicated.

The SMN complex was first linked to box C/D snoRNP biogenesis in a yeast two-hybrid screen, which indicated that the SMN protein interacts with the box C/D snoRNP common core protein fibrillarin (Liu and Dreyfuss, 1996; Jones et al, 2001). Further *in vitro* investigations revealed that SMN interacts with the RGG domain of fibrillarin (Pellizzoni et al, 2001; Jones et al, 2001). The fibrillarin interaction domain of SMN is not clear with data indicating the YG-box (Pellizzoni et al, 2001) and the Tudor domain (Jones et al, 2001; Whitehead et al, 2002). Immunoprecipitation assays from human cell extracts confirmed the fibrillarin and SMN interact *in vivo* (Pellizzoni et al, 2001; Jones et al, 2001). An interaction was also found between the SMN protein and the box H/ACA snoRNP common core protein GAR1 (Pellizzoni et al, 2001; Whitehead et al, 2002), which suggests that the SMN complex might be involved in the biogenesis of both box C/D and H/ACA snoRNPs.

The role of SMN in box C/D snoRNP biogenesis was further investigated by analysing the localisation of box C/D snoRNP factors in the absence of functional SMN. The expression of the dominant negative SMN, which was not functional in snRNP biogenesis, resulted in the accumulation of fibrillarin, the U3 box C/D snoRNA and the dominant negative SMN outside the nucleolus (Pellizzoni et al, 2001). Furthermore, in cells depleted of SMN the core proteins fibrillarin and NOP56 accumulated outside of the nucleolus in small nuclear bodies (Lemm et al, 2006). This indicates that SMN is required for the correct localisation of box C/D snoRNP factors, which suggests that SMN is required for box C/D snoRNP biogenesis.

The significance of the interaction of SMN with fibrillarin is not clear and while the SMN protein is required for the correct localisation of box C/D snoRNP components it is not known what effect this has on box C/D snoRNP biogenesis. Furthermore, it is not known whether other components of the SMN complex are involved in box C/D snoRNP biogenesis. To gain a greater understanding of the role of the SMN complex in box C/D snoRNP biogenesis the interactions of all the components of the SMN complex with the box C/D snoRNP common core proteins and assembly factors was characterised. A number of approaches were also used to determine whether the SMN protein interacts with the pre-snoRNP complex. In addition components of the SMN complex were depleted using siRNAs and the localisation and levels of box C/D snoRNAs analysed.

5.2. Results

5.2.1. Fibrillarin interacts with SMN but not with any of the other components of the SMN complex

It has been previously shown that the SMN protein binds to the RGG domain of fibrillarin (Pellizzoni et al, 2001; Jones et al, 2001); however, it was not known whether any of the other components of the SMN complex also interact with fibrillarin. To determine whether the SMN complex components Gemins2-8 or UNRIP interact with fibrillarin *in vitro* interaction assays were performed.

Fibrillarin and a fibrillarin deletion mutant (amino acids 82-321), which lacked the RGG containing domain, were expressed and purified from *E. coli* fused to N terminal GST tags (GST-fibrillarin and GST-N-C-fibrillarin; Figure 5.4 A; McKeegan et al, 2007). SMN, Gemins2-8 and UNRIP were expressed using reticulocyte lysate in the presence of [³⁵S] methionine to create *in vitro*-translated ³⁵S-labelled proteins (expression plasmids provided by U.Fischer, University of Wurtzburg, Germany). To perform the interaction assays GST-fibrillarin, GST-N-C-fibrillarin or GST alone (negative control) were immobilised to glutathione Sepharose beads and incubated with the *in vitro*-translated ³⁵S-labelled protein of interest. After two hours incubation at 4°C the glutathione Sepharose beads were washed, bound proteins separated on a 12 % SDS-PAGE gel and results visualised by autoradiography.

Immobilised GST-fibrillarin retained *in vitro*-translated ³⁵S-labelled SMN above the levels seen with the GST protein alone (Figure 5.4 B). This indicates, in agreement with previous studies (Pellizzoni et al, 2001; Jones et al, 2001), that fibrillarin interacts with the SMN. Immobilised GST-N-C-fibrillarin, which lacked the RGG domain, also retained the *in vitro*-translated ³⁵S-labelled SMN protein above the levels seen with GST alone. The retention of SMN by GST-N-C-fibrillarin, however, was lower than with full-length fibrillarin, which indicates that the RGG domain of fibrillarin is of importance for the interaction with SMN in agreement with previous studies (Pellizzoni et al, 2001; Jones et al, 2001).

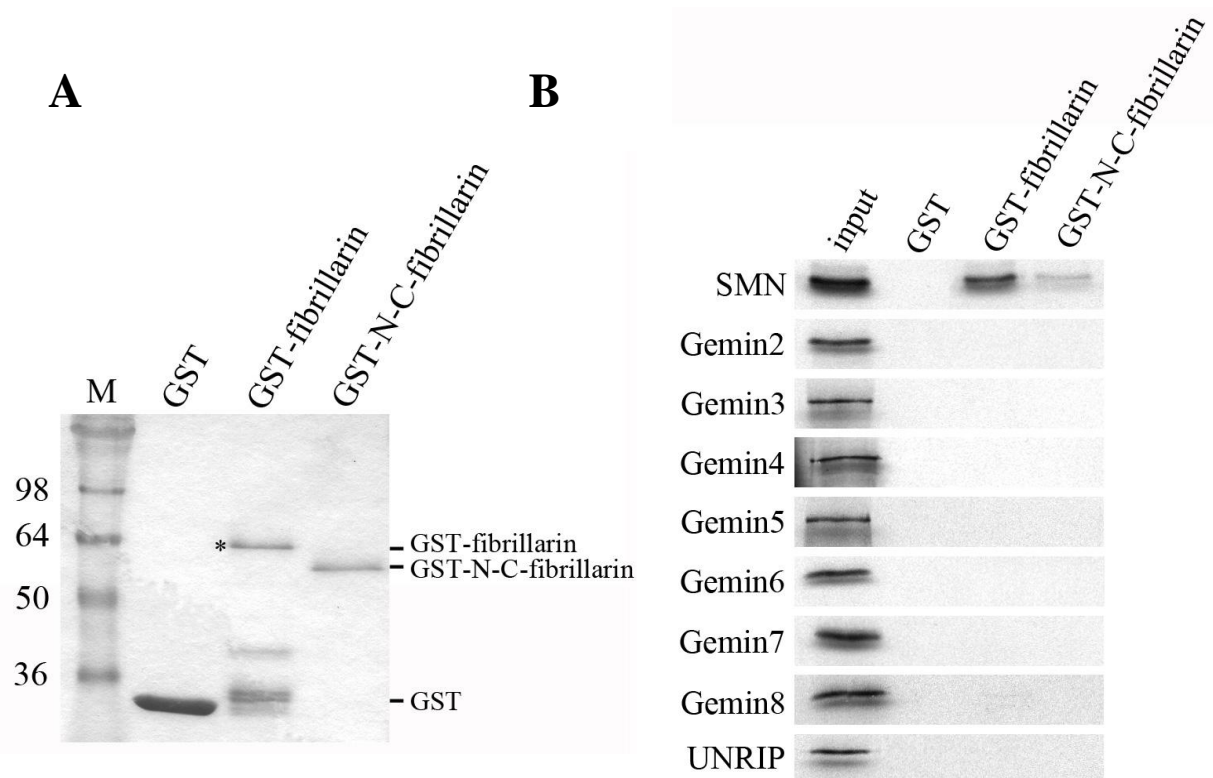


Figure 5.4: Interaction assay between fibrillarin and components of the SMN complex

(A) 12 % SDS-PAGE of GST, GST-fibrillarin and GST-N-C-fibrillarin visualised by Coomassie blue staining. The protein loaded is indicated at the top of each lane. M: molecular marker in kilodaltons. The migration of proteins is indicated to the right of the panel. Asterisk indicates the full-length protein where appropriate.

(B) Interaction assay of fibrillarin with SMN complex components. GST, GST-fibrillarin or GST-N-C-fibrillarin were immobilised to glutathione Sepharose beads and incubated with *in vitro*-translated ^{35}S -labelled SMN, Gemin2-8 or UNRIP. Bound proteins were separated on a 12 % SDS-PAGE gel and results visualised by autoradiography. Input represents 10 % of *in vitro*-translated ^{35}S -labelled protein used. The identity of GST tagged protein used is indicated above each lane. The identity of the radiolabelled protein used is indicated to the left of the panel. A single exposure was used for each *in vitro*-translated ^{35}S -labelled protein.

Immobilised GST-fibrillarin and GST-N-C-fibrillarin both failed to retain any of the *in vitro*-translated ^{35}S -labelled Gemin or the UNRIP proteins above the levels seen with GST alone (Figure 5.4 B). This indicates that fibrillarin does not interact with Gemin2-8 or UNRIP.

The ability of the other box C/D snoRNP common core proteins to interact with the SMN protein was also investigated. To determine whether 15.5K interacts with SMN

GST-15.K, expressed and purified from *E. coli* (McKeegan et al, 2007), was immobilised to glutathione Sepharose beads and incubated with *in vitro*-translated ³⁵S-labelled SMN. No interaction was seen between SMN and 15.5K (data not shown).

To determine if SMN interacts with NOP56 and NOP58 these common core proteins were expressed and purified from *E. coli* with an N terminal thioredoxin tag and C terminal His tag (His-NOP56, His-NOP58; McKeegan et al, 2007). Due to solubility issues deletion mutants of NOP56 (1-458) and NOP58 (1-435) had to be used; however, the C terminal charged regions that were removed have been shown not to be required for box C/D snoRNP formation in yeast (Gautier et al, 1997; Lafontaine and Tollervey, 2000). The His-NOP56 and His-NOP58 proteins were immobilised to protein A Sepharose beads, using anti-thioredoxin antibodies, incubated with *in vitro*-translated ³⁵S-labelled SMN and interactions visualised as described earlier. Immobilised His-NOP56 and His-NOP58 failed to retain *in vitro*-translated ³⁵S-labelled SMN above the levels of the control beads (data not shown). As His-NOP56 and His-NOP58 have been successfully used in other interaction assays (McKeegan et al, 2007; Figure 6.1) the lack of retention of *in vitro*-translated ³⁵S-labelled SMN indicates that NOP56 and NOP58 do not interact with SMN.

5.2.2. SMN interacts with the box C/D snoRNP assembly factors BCD1 and NUFIP

Box C/D snoRNP biogenesis proceeds in a large, multiprotein pre-snoRNP complex, which consists of snoRNA, common core proteins and assembly (BCD1, NOP17, NUFIP, TAF9, TIP48 and TIP49) RNA processing and transport factors (Watkins et al, 2004; 2007, Boulon et al, 2004; McKeegan et al, 2007). To determine whether the SMN complex interacts any of the box C/D snoRNP assembly factors, which could also indicate that it associates with the pre-snoRNP, *in vitro* interaction assays were performed.

The box C/D assembly factors NUFIP, NOP17, TIP48 and TIP49 were expressed and purified from *E. coli* fused to N terminal GST tags (GST-NUFIP, GST-NOP17, GST-TIP48 and GST-TIP49; Figure 4.7 A; McKeegan et al, 2007). There were issues with

producing recombinant full length BCD1; however, it was possible to efficiently express and purify a BCD1 deletion mutant (amino acids 1-360) with an N terminal GST tag and this was used in the interaction assay (GST-BCD1; McKeegan et al, 2007). The SMN complex components were expressed as *in vitro*-translated ^{35}S -labelled proteins as described earlier. The interaction assays were performed as described previously when using GST tagged and *in vitro*-translated ^{35}S -labelled proteins. As a positive control an interaction assay was performed between GST-fibrillarin and *in vitro*-translated ^{35}S -labelled SMN.

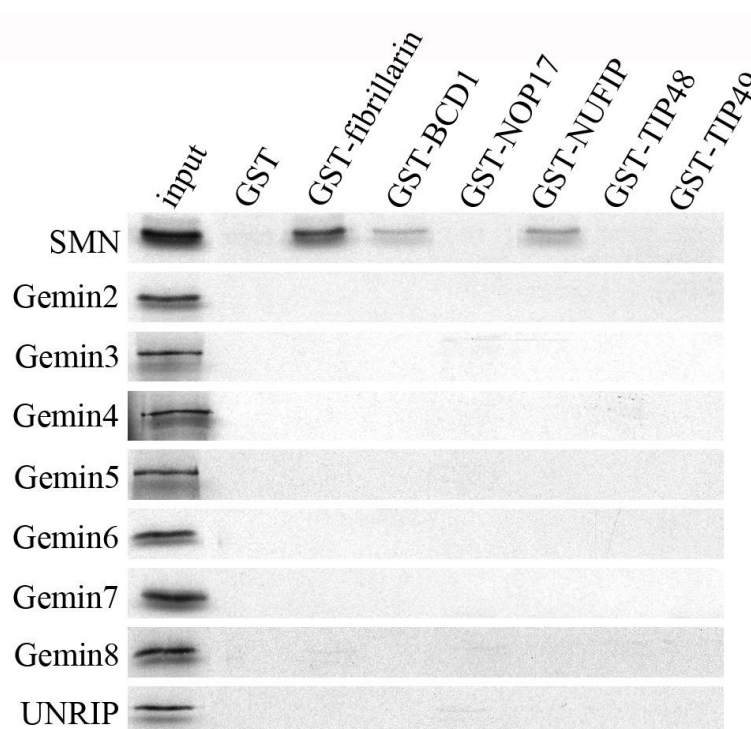


Figure 5.5: Interaction assay between components of the SMN complex and the box C/D snoRNP assembly factors
GST, GST-fibrillarin, GST-BCD1, GST-NOP17, GST-NUFIP, GST-TIP48 or GST-TIP49 were immobilised to glutathione Sepharose beads and incubated with *in vitro*-translated ^{35}S -labelled SMN, Gemins2-8 and UNRIP. Bound proteins were separated on an SDS-PAGE gel and results visualised by autoradiography. Input represents 10 % of *in vitro*-translated ^{35}S -labelled protein used. The identity of the GST tagged proteins used is indicated above each lane. The identity of the radiolabelled protein used is indicated to the left of the panel.

Immobilised GST-fibrillarin retained *in vitro*-translated ^{35}S -labelled SMN, but not any of the other SMN complex proteins, above the levels seen with GST alone (Figure 5.5). This indicated that SMN interacts with fibrillarin as has previously been shown (Figure

5.4). Immobilised GST-BCD1 and GST-NUFIP both retained *in vitro*-translated ³⁵S-labelled SMN above the levels seen when using GST alone; however, the retention of SMN was lower than with immobilised GST-fibrillarin. This indicates that the interaction of SMN with BCD1 and NUFIP was weaker than was seen with fibrillarin. Immobilised GST-BCD1 and GST-NUFIP failed to retain the *in vitro*-translated ³⁵S-labelled Gemins2-8 or UNRIP which indicates that no interactions occur between these proteins.

Immobilised GST-NOP17, GST-TIP48 and GST-TIP49 failed to retain *in vitro*-translated ³⁵S-labelled SMN, Gemins2-8 or UNRIP above the levels seen with GST alone (Figure 5.5). This suggests that NOP17, TIP48 and TIP49 do not interact with the SMN complex.

5.2.3. There is an overlap in the sedimentation of SMN, fibrillarin and the box C/D snoRNAs

It was not clear whether SMN interacts with the box C/D pre-snoRNP with previous studies providing contradictory results (Wehner et al, 2002; Watkins et al, 2004). To analyse the potential association of the SMN complex with the box C/D pre-snoRNP glycerol gradient centrifugation was performed. The co-migration of SMN and pre-snoRNP factors could suggest that they are present in the same complex.

HeLa SS6 nuclear extract was separated on a 10-30 % glycerol gradient and fractionated (fraction number one being the uppermost fraction of the gradient). RNA was extracted from each fraction, separated by electrophoresis on an 8 % acrylamide gel containing 7 M urea gel, Northern blot hybridisation performed using radiolabelled probes specific for the U3 and U8 box C/D snoRNAs and the results visualised by autoradiography. In addition proteins from each fraction were separated on a 12 % SDS-PAGE and analysed by Western blot assays using antibodies that recognise fibrillarin and SMN.

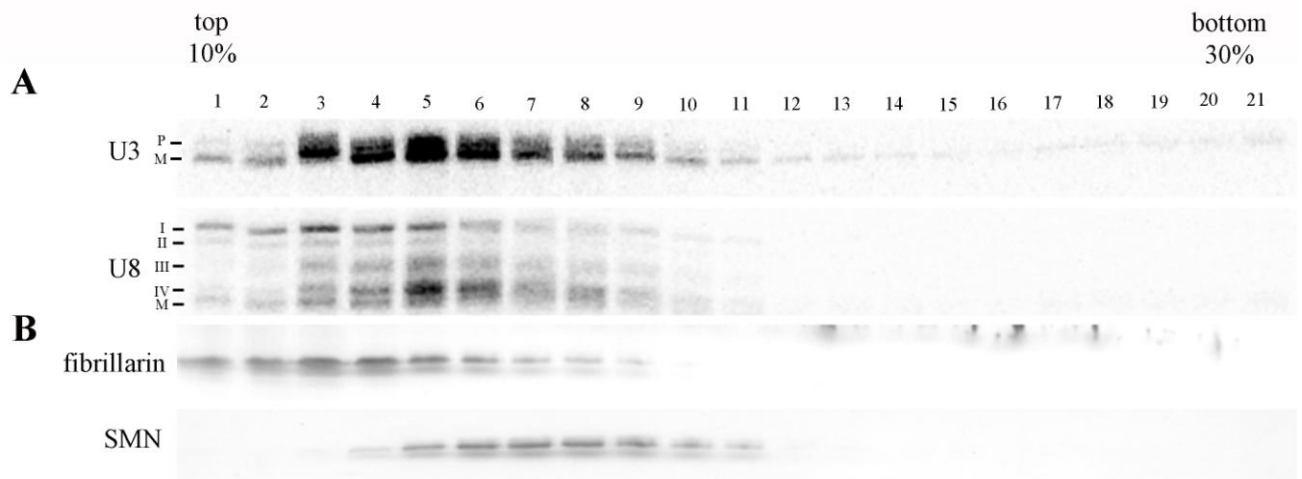


Figure 5.6: Glycerol gradient analysis of HeLa SS6 nuclear extract

(A) HeLa SS6 nuclear extract was separated on a 10-30 % glycerol gradient. RNA was extracted from individual fractions, separated on an 8 % acrylamide gel containing 7 M urea and analysed by Northern blot hybridisation using radiolabelled probes specific for the U3 and U8 box C/D snoRNAs. Results were visualised by autoradiography. The position of U8 precursors (I to IV) and U3 precursors (P) as well as mature length snoRNAs (M) are indicated to the left of the panel (B) Proteins from individual fractions was separated on a 12 % SDS-PAGE and Western blot analysis performed using antibodies that recognise fibrillarin and SMN (indicated left of the panel).

Both precursor and mature U3 and U8 box C/D snoRNAs are visible on the glycerol gradient and exhibit different sedimentation behaviours (Figure 5.6 A). The precursor and mature U3 snoRNA sedimented in a broad peak, between fractions 1 to 9, with the majority of mature U3 snoRNA in fraction 5. Unlike the U3 snoRNA, which only has one length of precursor, there are numerous intermediate U8 precursors (I to IV), which indicate a more complex 3' RNA processing pathway than the U3 snoRNA (Watkins et al, 2007). The U8 snoRNA sedimented in a broad peak between fractions 1 to 9; however, the two longest U8 intermediates (I and II) were found mostly in fractions 1 to 5, while the U8 intermediate III precursors were predominately in fractions 3 to 6. The shortest U8 intermediate (IV) was mostly found in fractions 3 to 6 while the mature length U8 snoRNA was predominately in fractions 3 and 4.

The box C/D snoRNP common core protein fibrillarin co-sedimented with both U3 and U8 snoRNAs in fractions 1 to 9 and was predominately found in fractions 3 and 4 (Figure 5.6 B). As fractions 3 and 4 contain mature U8 snoRNA and fibrillarin these fractions, most likely, represent mature U8 snoRNPs. Fraction 5 contains both fibrillarin

and mature U3 snoRNA and therefore, more than likely, represents mature U3 snoRNPs.

The SMN protein was present in fractions 4 to 9 but predominately in fractions 6, 7 and 8 (Figure 5.6 B). SMN did not co-sediment with the box C/D snoRNA and / or fibrillarin; however, there was some degree of overlap between fractions 4 and 10. It is therefore possible that SMN, fibrillarin and the box C/D snoRNA are in the same complex; however, this cannot be deduced by this technique alone.

5.2.4. SMN is not a stable component of the U3 and U8 box C/D pre-snoRNPs

SMN has been shown to interact with the box C/D snoRNP common core protein fibrillarin both *in vitro* and *in vivo* (Figure 5.4; Pellizzoni et al, 2001; Jones et al, 2001). It is, however, not clear whether SMN associates with the box C/D pre-snoRNP (Wehner et al, 2002; Watkins et al, 2004). To determine whether SMN interacts with the box C/D pre-snoRNP and also to confirm the *in vivo* interaction of SMN and fibrillarin immunoprecipitation assays were performed on extracts produced from cells that express FLAG tagged SMN or fibrillarin.

Whole cell, nucleoplasmic and nucleolar extracts were produced from induced FLAG-SMN, FLAG-fibrillarin and control HEK293 cell lines. The control HEK293 cell line provides a negative control as it only expresses the FLAG and His tags. Anti-FLAG antibodies were immobilised onto protein A Sepharose beads and incubated with the cell extract of interest. The protein A Sepharose beads were then washed and co-precipitated RNA extracted. The RNA samples were separated by electrophoresis on an 8 % acrylamide gel containing 7 M urea, Northern blot hybridisation performed using a radiolabelled probes specific for the U3 and U8 box C/D snoRNAs and the spliceosomal U1 snRNA, and the results visualised by autoradiography. In addition co-precipitated proteins were separated on a 12 % SDS-PAGE gel and Western blot analysis performed using antibodies that recognise fibrillarin. As FLAG-fibrillarin migrates significantly slower than endogenous fibrillarin on SDS-PAGE gels both forms are distinguishable using only anti-fibrillarin antibodies.

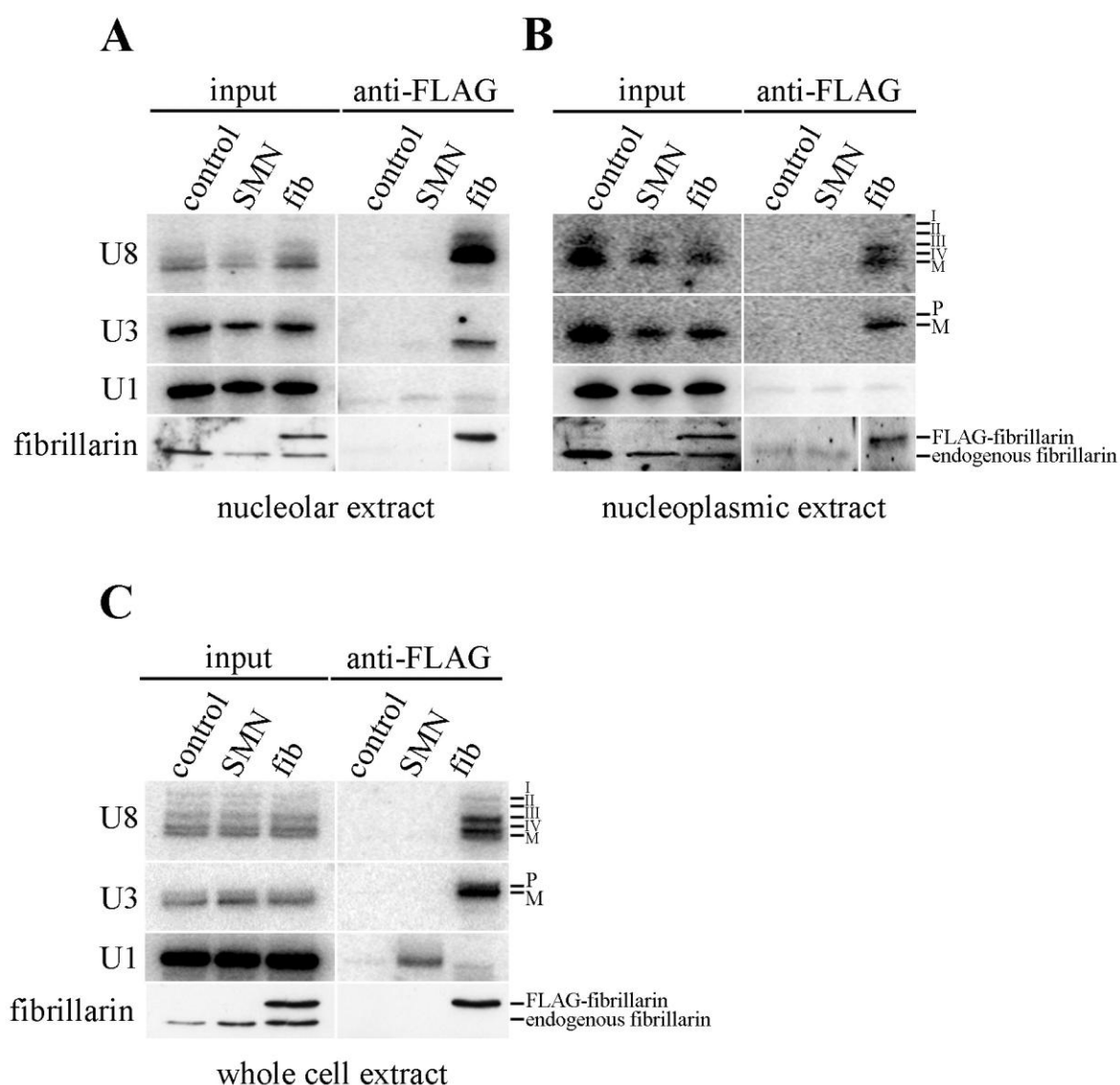


Figure 5.7: Immunoprecipitation of FLAG-SMN and FLAG-fibrillarin from HEK293 cell extracts

(A) Nucleolar, (B) nucleoplasmic and (C) whole cell extracts were produced from HEK293 cells that expressed FLAG tagged fibrillarin (fib), SMN or just the FLAG tag (control). Immunoprecipitation assays were performed using anti-FLAG antibodies and co-precipitating RNA extracted, separated by electrophoresis on an 8 % acrylamide gel containing 7 M urea and analysed by Northern blot hybridisation using radiolabelled probes specific for the U3 and U8 box C/D snoRNAs and the spliceosomal U1 snRNA. Results were visualised by autoradiography. Co-precipitating proteins were separated on a 12 % SDS-PAGE gel and analysed by Western blot assays using antibodies that recognised fibrillarin. The cell lines used are indicated above each lane. Input represents 10 % of the amount of extract used in immunoprecipitation assay. The factors probed for are indicated to the left of the panels. The migration of U8 precursors (I to IV) and the U3 precursor (P) as well as mature length snoRNAs (M) is indicated to the right of the panels. The migration of endogenous and FLAG-fibrillarin are shown to the right of the panels.

Anti-FLAG antibodies successfully co-precipitated both the U3 and U8 box C/D snoRNAs from all the FLAG-fibrillarin HEK293 cell extracts but not from any of the control HEK293 cell extracts (Figure 5.7 A, B and C). This indicates that FLAG-fibrillarin associated with box C/D snoRNAs in both the nucleoplasm and nucleolus. As mature box C/D snoRNPs are found in the nucleolus (Verheggen et al, 2001) the co-precipitation of the U3 and U8 snoRNA from FLAG-fibrillarin nucleolar extracts (Figure 5.7 A) suggests that the FLAG-fibrillarin was incorporated into mature box C/D snoRNPs. The co-precipitation of U3 and U8 pre-snoRNAs from the FLAG-fibrillarin cell extracts also indicates that FLAG-fibrillarin associated with the box C/D snoRNPs before mature length snoRNA was produced (Figure 5.7 C).

Interestingly anti-FLAG antibodies did not co-precipitate endogenous fibrillarin from any of the FLAG-fibrillarin HEK293 cell extracts (Figure 5.7 A, B and C). This was surprising, as box C/D snoRNPs have been predicted to contain two copies of fibrillarin (Cahill et al, 2002; Aittaleb et al, 2003) and therefore the co-precipitation of endogenous fibrillarin, associated in the same box C/D snoRNP complex as the FLAG-fibrillarin, was expected. The lack of co-precipitation of endogenous fibrillarin could indicate that box C/D snoRNPs only contain one copy of fibrillarin.

Anti-FLAG antibodies successfully co-precipitated U1 snRNA from FLAG-SMN HEK293 whole cell extract, but not from the control whole cell extract (Figure 5.7 C). In addition, the anti-FLAG antibodies did not co-precipitate U1 snRNA from either the FLAG-SMN HEK293 nucleoplasmic or nucleolar extracts (Figure 5.7 A and B). This indicates that FLAG-SMN interacted with the U1 snRNA specifically in the cytoplasm, which was in agreement with SMN facilitating the association of the Sm core with snRNA in the cytoplasm (Pellizzoni et al, 2002; Yong et al, 2002, 2004; Wehner et al, 2002). Anti-FLAG antibodies did not co-precipitate the U3 or U8 box C/D snoRNAs from any of the FLAG-SMN HEK293 cell extracts (Figure 5.7 A, B and C). This indicated that FLAG-SMN did not stably associate with the box C/D pre-snoRNP. Surprisingly anti-FLAG antibodies did not co-precipitate fibrillarin from any of the FLAG-SMN HEK293 cell extracts, which indicated that FLAG-SMN did not interact with fibrillarin. This is not in agreement with previous studies that show fibrillarin can be co-precipitated using anti-SMN antibodies (Pellizzoni et al, 2001; Jones et al, 2001).

5.2.5. Gemin5 is essential for the accumulation of box C/D snoRNA

Previous studies have shown that in the absence of the box C/D snoRNP assembly factors there is a reduction of box C/D snoRNA levels, most likely, due to a decrease in box C/D snoRNP biogenesis (Peng et al, 2003; Watkins et al, 2004, 2007; McKeegan et al, 2007; Boulon et al, 2008). To determine whether the SMN complex is essential for box C/D snoRNP biogenesis SMN and the Gemin proteins were depleted using siRNAs and the levels of box C/D snoRNA analysed. It was, however, first necessary to assess the efficiency of the siRNA duplexes.

The siRNAs that were used to target SMN, Gemin2, 5, 6 and 7 have all been shown to deplete the levels of the respective mRNAs (data not shown, control experiments performed by I.Lemm and R.Luhrmann, Max Planck Institute for Biophysical Chemistry, Germany). Unfortunately, it was not possible to obtain siRNA duplexes which efficiently depleted the levels of Gemin3, 4, 8 or UNRIP. To confirm the efficiency of the siRNAs with the transfection protocol used in this study HeLa SS6 cells were transfected with siRNAs targeting SMN or Gemin2 mRNA. These two siRNAs were selected as antibodies were not available for the other SMN complex proteins. As a control, HeLa SS6 cells were transfected with functional siRNAs that targeted firefly luciferase mRNA, which is not present in human cells (Elbashir et al, 2002). After sixty hours cells were harvested, proteins were separated on a 12 % SDS-PAGE gel and Western blot analysis performed using antibodies that recognise SMN, Gemin2 and, as a control, fibrillarin.

In cells transfected with siRNAs targeting SMN there was dramatic depletion of SMN protein levels compared to the control cells (Figure 5.8). The level of fibrillarin was unaffected in cells transfected with the SMN siRNA, which indicates that the depletion of SMN was specific and also that the levels of SMN do not effect fibrillarin levels. In cells transfected with siRNAs targeting Gemin2 there was a significant reduction in Gemin2 protein levels compared to the control cells. The level of SMN was unaffected by the depletion of Gemin2 which indicates that not only were the Gemin2 siRNAs specific but that the loss of Gemin2 does not affect the level of SMN.

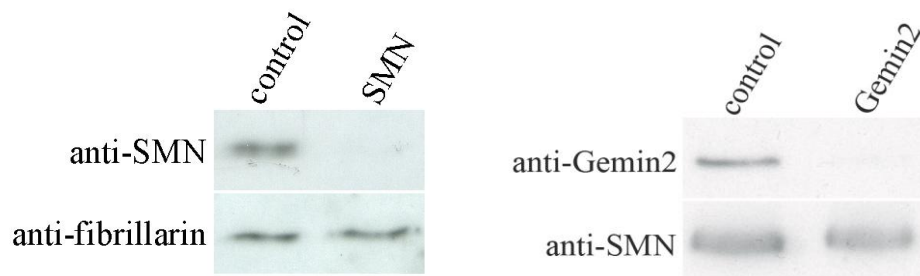


Figure 5.8: Depletion of SMN and Gemin2

HeLa SS6 cells were transfected with siRNAs targeting either firefly luciferase (control), SMN or Gemin2 mRNA. After sixty hours cells were harvested, proteins separated on a 12 % SDS-PAGE gel and Western blot assays performed using antibodies that recognised SMN, fibrillarin or Gemin2. The siRNA duplexes used are indicated above each panel. Antibodies used in Western blot assays are indicated to the left of the panels.

To determine whether the SMN complex is required for box C/D snoRNP biogenesis HeLa SS6 cells were transfected with siRNAs targeting firefly luciferase (control), SMN, Gemin2, 5, 6 and 7. As a positive control siRNAs targeting NOP58, which is essential for box C/D snoRNA accumulation, were also transfected (Watkins et al, 2004). After sixty hours cells were harvested and RNA (derived from an equal numbers of cells) separated by electrophoresis on a 8 % acrylamide gel containing 7 M urea. Northern blot hybridisation was then performed using radiolabelled probes specific for the U3, U8, U14 box C/D snoRNAs, the spliceosomal U1 snRNA and the ncRNA component of SRP, 7SL. 7SL was chosen as a internal control of RNA levels as it has no known function with any of the proteins targeted for depletion. The results were visualised by autoradiography (Figure 5.9 A), quantified, adjusted based upon the 7SL levels (which were unaffected by any of the siRNA depletions) and represented as a percentage change compared to cells transfected with the control siRNA (Figure 5.9 B).

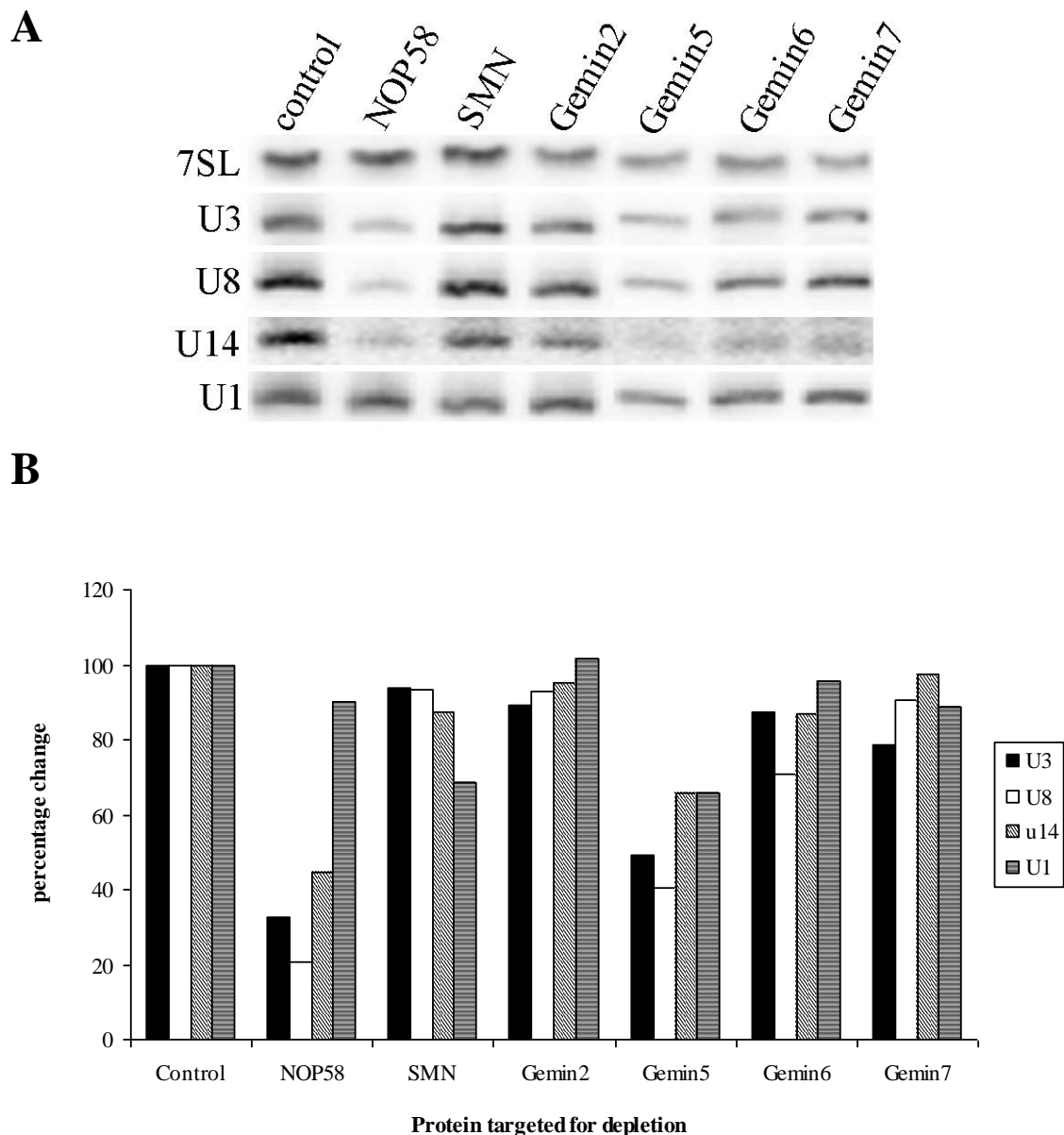


Figure 5.9: Box C/D snoRNA levels after depletion of components of the SMN complex

HeLa SS6 cells were transfected with siRNAs targeting either firefly luciferase (control), NOP58, SMN, Gemin2, 5, 6 or 7. After sixty hours cells were harvested, RNA extracted, separated by electrophoresis on 8 % acrylamide gel containing 7 M urea and Northern blot hybridisation performed using radiolabelled probes specific for the U3, U8, U14 box C/D snoRNAs, the U1 snRNA and 7SL. Results were visualised by autoradiography. (A) Northern blot with protein targeted for depletion indicated above the panel. The radiolabelled RNA probes used are indicated to the left of each panel. (B) Quantification of RNA levels of Figure 5.9 A. The Levels of box C/D snoRNA and snRNA were all adjusted relative to the level of the internal control, 7SL, and represented as a percentage change.

In cells depleted of NOP58 there was a dramatic reduction in the levels of U3 box C/D snoRNA to 32 % of that in the control cells (Figure 5.9 B). Similar reductions were seen for all the box C/D snoRNAs in cells depleted of NOP58, which was in agreement with previous studies that have shown that NOP58 is essential for box C/D snoRNA accumulation (Lafontaine and Tollervy, 2000; Watkins et al, 2004). The depletion of NOP58 did not result in a change in the levels of U1 snRNA, which was expected, as NOP58 is not known to be involved in snRNP biogenesis.

Depletion of Gemin5 resulted in the reduction of all the box C/D snoRNAs; U3 levels were reduced to 49 %, U8 levels to 40 % and U14 levels to 65 % of that in the control cells (Figure 5.9 B). This indicates that Gemin5 was essential for box C/D snoRNA accumulation and suggests that it may be involved in box C/D snoRNP biogenesis. In cells depleted of Gemin5 the U1 snRNA levels were reduced to 65 %, compared to the control cells, which indicates that Gemin5 was essential for snRNA accumulation and suggests it functions in snRNP biogenesis.

The loss of SMN did not result in a significant change in the levels of the box C/D snoRNAs (Figure 5.9 B). This indicates that SMN was not essential for box C/D snoRNA accumulation. In cells depleted of SMN there was a reduction in the levels of U1 snRNA to 68 % of that in the control cells. This showed that SMN was required for snRNA accumulation, which was expected as SMN has a known role in the biogenesis of snRNPs (Pellizzoni et al, 2002; Wehner et al, 2002; Yong et al, 2002; 2004). In cells depleted of Gemin2, 6 and 7 there was no change in the levels of any of the box C/D snoRNAs or U1 snRNA (Figure 5.9 B), showing that these proteins were not required for their accumulation.

The results presented here show that out of the SMN complex only Gemin5 was essential for box C/D snoRNA accumulation and could indicate that it is involved in box C/D snoRNP biogenesis.

5.2.6. The SMN complex is essential for the correct localisation of box C/D snoRNA

Previous studies have shown that expression of non-functional SMN effects the localisation of box C/D snoRNA and fibrillarin (Pellizzoni et al, 2001). Furthermore, the depletion of SMN results in the mislocalisation of the box C/D snoRNP common core proteins fibrillarin and NOP56 (Lemm et al, 2006). As the localisation of box C/D snoRNP factors is an indication of biogenesis SMN, Gemin2, 5, 6 and 7 were depleted and the localisation of box C/D snoRNA analysed.

HeLa SS6 cells, grown on coverslips, were transfected with siRNAs targeting firefly luciferase (control), SMN, Gemin2, 5, 6 or 7. After sixty hours cells were fixed and FISH performed using fluorescent probes specific for either the U3 or U8 box C/D snoRNAs, or the spliceosomal U2 snRNA. The U2 snRNA provided a marker of the nucleoplasmic Cajal bodies and the nuclear speckles. In addition, as depletion of SMN results in the loss of the U2 snRNA from Cajal bodies (Shpargel and Matera, 2005; Lemm et al, 2006) it provided a positive control for the siRNA mediated depletion of SMN.

In the control cells the U3 box C/D snoRNA was found in Cajal bodies but predominately in the nucleolus (Figure 5.10 A), the typical distribution of U3 snoRNA (Figure 5.3). In cells depleted of SMN, while the U3 snoRNA was still present in the nucleolus, there was an accumulation in the nucleoplasm and also in the cytoplasm. There was also a loss of the U3 snoRNA localised to Cajal bodies; however, it is difficult to interpret this as in SMN depleted cells there is a reduction in Cajal body numbers (Shpargel and Matera, 2005; Lemm et al, 2006).

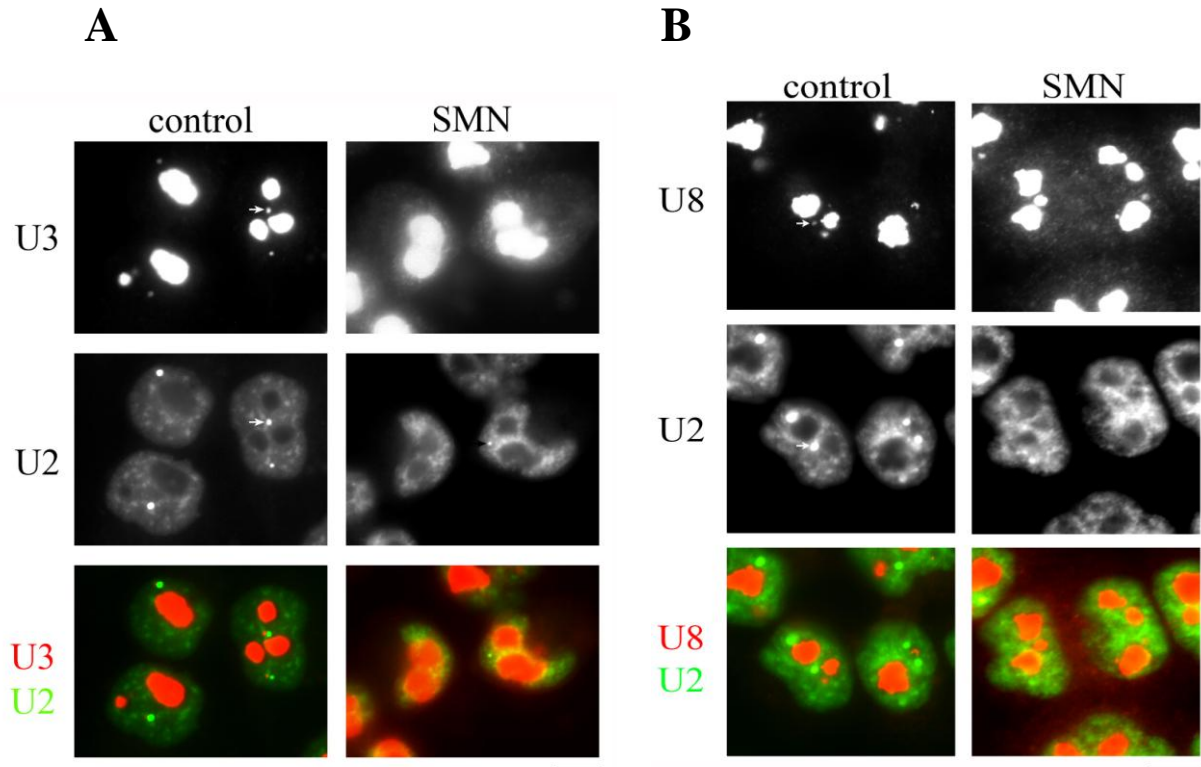


Figure 5.10: Box C/D snoRNA localisation in SMN depleted cells

HeLa SS6 cells were transfected with siRNAs targeting either firefly luciferase (control) or SMN. After sixty hours cells were fixed and FISH performed using florescent probes specific for the (A) U3 or (B) U8 box C/D snoRNAs, and the U2 snRNA. The same exposure time was used for each probe to allow direct comparison of RNA levels and distribution. The protein targeted for depletion is indicated at the top of each column of images. The labels along the left hand side of the panels indicate the FISH probes used. The bottom row shows a merged image of U3 or U8 snoRNA (red) and U2 snRNA (green). The white arrows show a selected Cajal body. The bar to the bottom right of panels represents 5 µm. Figure 5.10 B provided by N.J.Watkins, Newcastle University, UK.

To determine whether SMN was required for the localisation of other box C/D snoRNAs the U8 box C/D snoRNA was also analysed in SMN depleted cells (Figure 5.10 B). In the control cells the U8 snoRNA was predominately localised to the nucleolus but also found in the Cajal bodies. In the SMN depleted cells the U8 snoRNA accumulated in the nucleoplasm, cytoplasm and was lost from Cajal bodies, as was seen for the U3 box C/D snoRNA (Figure 5.10 A). This indicates that SMN is required for the localisation of box C/D snoRNAs and is involved in box C/D snoRNP biogenesis.

In cells depleted of Gemin2 the U3 box C/D snoRNA was predominately found in the nucleolus; however, there was an accumulation in nucleoplasm (Figure 5.11). Unlike cells depleted of SMN no U3 box C/D snoRNA accumulated in the cytoplasm and both the U3 snoRNA and U2 snRNA were found in Cajal bodies. In cells depleted of Gemin5 the U3 snoRNA was again found in the nucleolus, however, there was an accumulation in both the nucleoplasm and cytoplasm. Also with the depletion of Gemin5 there was a loss of both the U3 snoRNA and U2 snRNA from Cajal bodies. This could indicate that either Gemin5 is essential for localisation of U3 snoRNA and U2 snRNA to the Cajal bodies or that Gemin5 is required for the integrity of Cajal bodies, as is the case for SMN (Shpargel and Matera, 2005; Lemm et al, 2006). The depletion of Gemin6 resulted in the accumulation of the U3 snoRNA outside the nucleolus in the nucleoplasm and cytoplasm and the loss of the U3 snoRNA and U2 snRNA from Cajal bodies. In cells depleted of Gemin7 the U3 snoRNA was found in the nucleolus but also accumulated in the nucleoplasm. The depletion of Gemin7 also caused a loss of U3 snoRNA and U2 snRNA from Cajal bodies.

The loss of any of the Gemin proteins tested resulted in an accumulation of U3 box C/D snoRNA in the nucleoplasm and in some cases the cytoplasm. This indicates that the Gemin2, 5, 6 and 7 are required for the correct localisation of box C/D snoRNA, which suggests that they are involved in box C/D snoRNP biogenesis. The Loss of Gemin5, 6 and 7 also resulted in the loss of U3 box C/D snoRNA from Cajal bodies; however, as the U2 snRNA was also lost from Cajal bodies it is difficult to interpret these results.

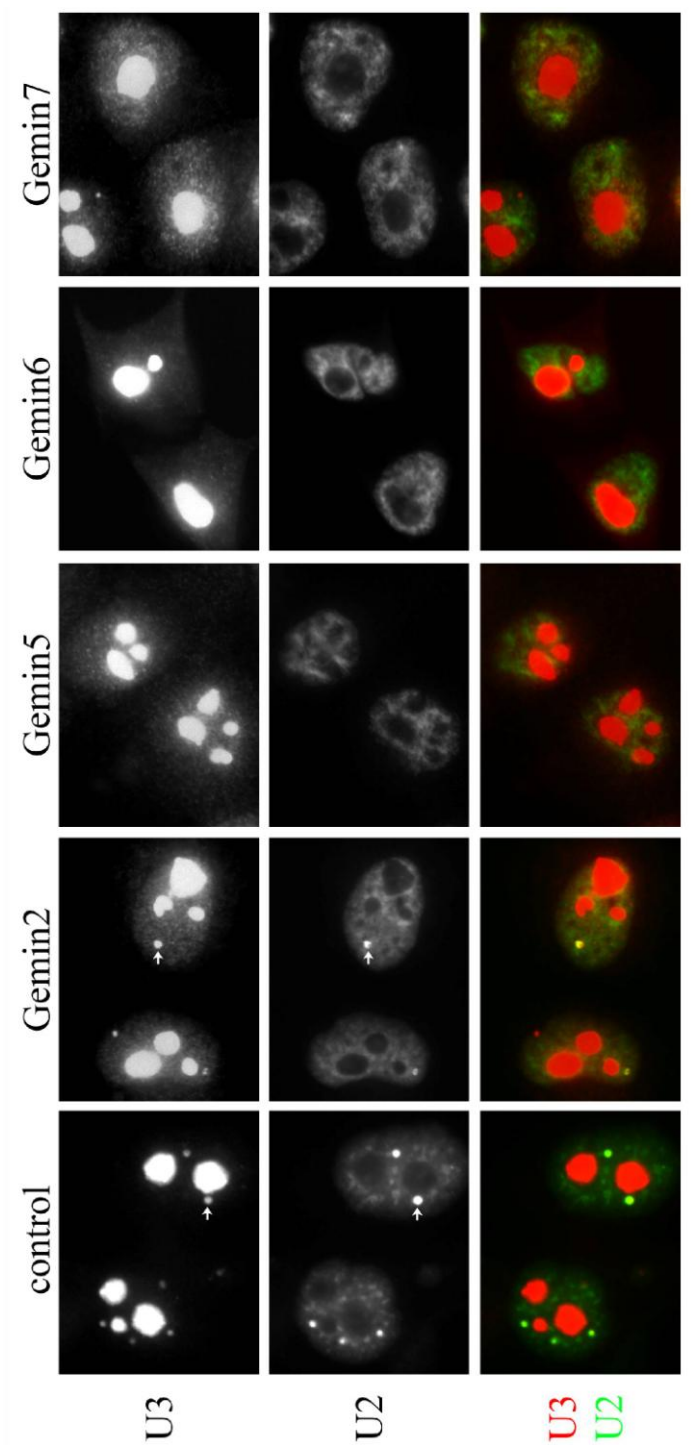


Figure 5.11: The Localisation of Box C/D snoRNAs in cells depleted of Gemin2, 5, 6 and 7

HeLa SS6 cells were transfected with siRNAs targeting either firefly luciferase (control) or Gemin2, 5, 6 and 7. Cells were fixed after sixty hours and analysed by FISH using fluorescent probes specific for the U3 box C/D snoRNA and U2 snRNA. The same exposure time was used for each probe to allow direct comparison of RNA levels and distribution. The protein targeted for depletion is indicated at the top of each column of images. The bottom row shows a merged image of U3 or U2 snRNA (red) and U2 snRNA (green). The white arrows show a selected Cajal body. The bar to the bottom right of panel represents 5 μ m.

To ensure that the accumulation of box C/D snoRNA in the nucleoplasm and cytoplasm was due to specific effects of SMN depletion and not due to a general cell defect caused by the loss of an essential factor the localisation of the unrelated 7SL RNA was analysed in cells depleted of SMN. As a comparison the localisation of 7SL was also analysed in cells depleted of the box C/D core protein NOP58.

In the control cells 7SL was found to be predominately in the cytoplasm and to a lesser extent the nucleoplasm (Figure 5.12). 7SL was not found in the nucleolus of the control cells, which were indicated by the U3 box C/D snoRNA. In the cells depleted of NOP58 there was a dramatic reduction in the levels of the U3 snoRNA, as shown earlier (Figure 5.9 B), but the level and distribution of 7SL was unchanged (Figure 5.12). This indicates that the depletion of NOP58 specifically affected the U3 snoRNA. In the SMN depleted cells the U3 snoRNA accumulated in the nucleoplasm and cytoplasm; however, the level and distribution of 7SL was unaffected. This suggests that the accumulation of box C/D snoRNA in the nucleoplasm and cytoplasm in SMN depleted cells was due to the loss of box C/D snoRNP specific functions of SMN.

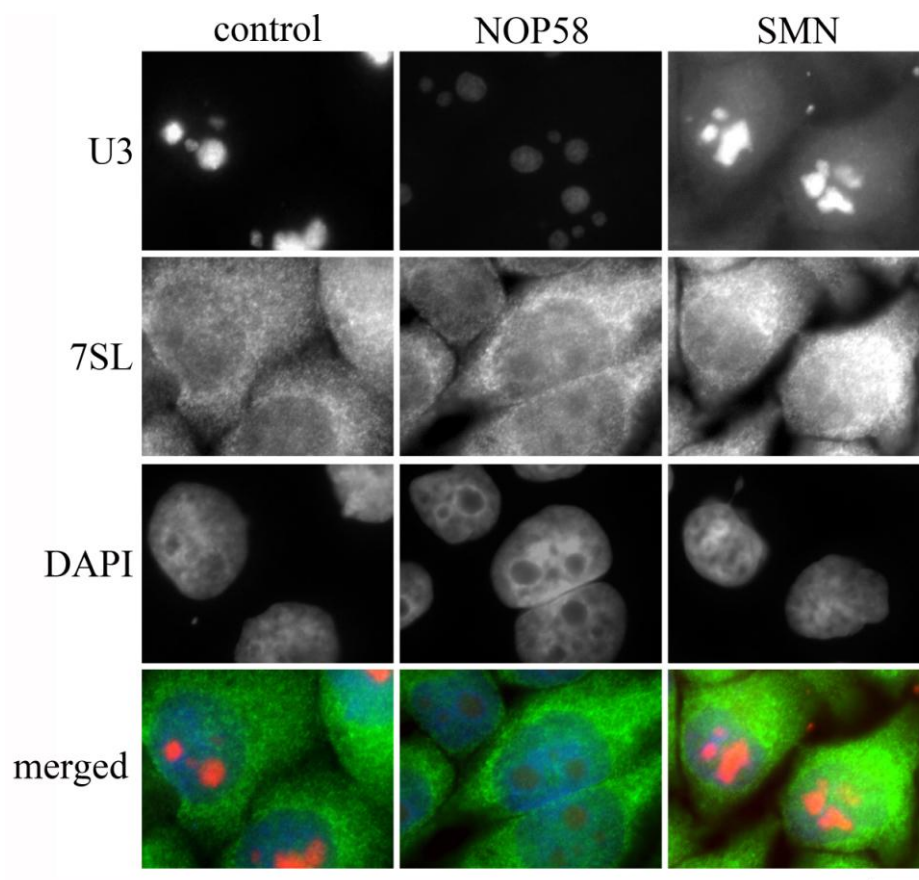


Figure 5.12: The localisation of 7SL in SMN depleted cells

HeLa SS6 cells were transfected with siRNA targeting either firefly luciferase (control), NOP58 or SMN. Cells were fixed after sixty hours and analysed by FISH using fluorescent probes specific for the U3 box C/D snoRNA and 7SL. DAPI staining was used to show the nucleoplasm. The same exposure time was used for each probe to allow direct comparison of RNA levels and distribution. The protein targeted for depletion is indicated at the top of each column of images. The labels along the left hand side of the panels indicate the FISH probes used and DAPI staining of the nucleoplasm. The bottom row shows a merged image of the U3 snoRNA (red), with 7SL (green) and DAPI (blue). The bar to the bottom right of the panel represents 5 μ m.

5.3. Discussion

The SMN protein has been implicated in box C/D snoRNP biogenesis in a number of studies (Pellizzoni et al, 2001; Jones et al, 2001; Whitehead et al, 2002; Lemm et al, 2006); however, the precise role of SMN and whether it functions in box C/D snoRNP biogenesis as part of the SMN complex is not clear.

Previous studies using Yeast-2-hybrid screens, *in vitro* interaction assays and immunoprecipitation assays have shown that the box C/D snoRNP common core protein fibrillarin interacts with the SMN protein. The N terminal RGG domain of fibrillarin mediates the interaction with SMN; however, the interaction domain of SMN is not clear (Pellizzoni et al, 2001; Jones et al, 2001). In this study *in vitro* interaction assays were performed between fibrillarin and all the components of the SMN complex to determine if other interactions existed. These assays revealed that fibrillarin only interacts with the SMN protein and that the RGG domain of fibrillarin is important for this interaction (Figure 5.4). Further *in vitro* interaction assays showed that there were no interactions between the SMN protein and the other box C/D snoRNP common core proteins. The interaction between fibrillarin and SMN was also investigated *in vivo* using immunoprecipitation assays, utilising extract from cells that expressed FLAG-SMN (Figure 5.7). It was, however, not possible to co-precipitate fibrillarin with anti-FLAG antibodies. The lack of a detectable *in vivo* interaction suggests that weak and / or transient interactions occur between SMN and fibrillarin.

The biogenesis of box C/D snoRNPs occurs in a dynamic, multiprotein complex known as the pre-snoRNP, which consists of snoRNA, common core proteins and assembly, transport and RNA processing factors (Watkins et al, 2004; McKeegan et al, 2007). To determine whether the SMN complex interacts with the pre-snoRNP a number of approaches were utilised. *In vitro* interaction assays were performed between the box C/D snoRNP assembly factors and components of the SMN complex. These assays showed that the SMN protein interacts with the box C/D snoRNP assembly factors BCD1 and NUFIP (Figure 5.5), which could indicate that SMN associates with the pre-snoRNP. Another approach used to determine if SMN interacts with the pre-snoRNP was to separate HeLa SS6 nuclear extract on a glycerol gradient and analyse the

sedimentation of SMN and components of the box C/D snoRNPs (Figure 5.6). The glycerol gradient revealed that there was some degree of overlap in sedimentation of SMN, fibrillarin and box C/D snoRNA, which could indicate that SMN interacts with the box C/D pre-snoRNP. Immunoprecipitation assays were also performed to determine whether SMN interacts with the box C/D pre-snoRNP. As previous studies have produced conflicting data regarding the co-precipitation of box C/D snoRNA with anti-SMN antibodies (Wehner et al, 2002; Watkins et al, 2004) immunoprecipitation experiments were performed using anti-FLAG antibodies and extracts from cells that expressed FLAG-SMN (Figure 5.7). No co-precipitation of box C/D snoRNA with anti-FLAG antibodies was detected, indicating that FLAG-SMN did not interact with the box C/D pre-snoRNP. Overall the data suggests that if the SMN complex does interact with the box C/D pre-snoRNP, as is indicated by the *in vitro* interaction of SMN with fibrillarin, BCD1 and NUFIP, then it is not a stable component.

To determine whether the SMN complex is required for box C/D snoRNP biogenesis SMN, Gemin2, 5, 6 and 7 were depleted using siRNAs and the level of box C/D snoRNA analysed. Unfortunately, it was not possible to efficiently deplete Gemin3, 4 or UNRIP and so these proteins were not analysed. Of all the SMN complex factors analysed only the depletion of Gemin5 resulted in the reduction in the levels of box C/D snoRNA (Figure 5.10). As Gemin5 is essential for the accumulation of box C/D snoRNA this indicates that it may be involved in box C/D snoRNP biogenesis. Gemin5 was not shown to interact with any of the box C/D snoRNP factors tested; however, Gemin5 has shown to interact directly with the snRNAs (Battle et al, 2006, 2007). It is therefore possible that the RNA binding capabilities of Gemin5 could extend to box C/D snoRNAs. Gemin5 could be essential for box C/D snoRNA accumulation through functions that are required for the stability of RNA, such as the association of common core proteins or direct binding of Gemin5 to the box C/D snoRNA preventing degradation. Alternatively the loss of Gemin5 could result in a decrease in the rate of box C/D snoRNP biogenesis leading to a reduction in box C/D snoRNA levels.

The level of the U1 snRNA was also analysed in cells depleted of the SMN, Gemin2, 5, 6 and 7 (Figure 5.9). Only the depletion of SMN and Gemin5 resulted in a reduction in snRNA levels, which shows that these proteins are essential for snRNA accumulation and suggests that they are involved in snRNP biogenesis. The SMN complex has a defined role in snRNP biogenesis; however, the extent to which the different components function in snRNP biogenesis is not clear. Recently a minimal complex consisting of just SMN and Gemin2 has been shown to be sufficient for snRNP biogenesis (Kroiss et al, 2008; Chari et al, 2008). This could imply that the other Gemin proteins have redundant functions or are involved in processes outside of snRNP biogenesis.

Another approach used to analyse the role of the SMN complex in box C/D snoRNP biogenesis was to determine whether it was essential for the localisation of box C/D snoRNPs. Previous studies have shown that in the absence of functional SMN there is an accumulation of box C/D snoRNP factors outside the nucleolus; however, the importance of the Gemin proteins was not addressed (Pellizoni et al, 2001; Lemm et al, 2006). In cells depleted of SMN there was an accumulation of both the U3 and U8 box C/D snoRNAs in the nucleoplasm and also the cytoplasm (Figure 5.10). In cells depleted of Gemin5 and 6 the U3 snoRNA accumulated in the nucleoplasm and the cytoplasm of cells (Figure 5.11). In cells depleted of Gemin2 and 7 the U3 snoRNA accumulated in the nucleoplasm. Together this data indicates that SMN and Gemin2, 5, 6 and 7 are vital for the correct localisation of box C/D snoRNA and suggests that all these factors are involved in box C/D snoRNP biogenesis. A possible explanation for the accumulation of the U3 box C/D snoRNA in the nucleoplasm and cytoplasm is that in the absence of SMN and Gemin2, 5, 6 and 7 there is a block in box C/D snoRNP biogenesis resulting in the formation of pre-snoRNP complexes which cannot localise to the nucleolus. This could be due to the SMN complex functioning as a transport factor or the SMN complex is involved in the recruitment and / or association of common core proteins, which is essential for nucleolar localisation (Figure 3.4; Verheggen et al, 2001, 2002).

The depletion of SMN, Gemin5, 6 and 7 also resulted in the loss of box C/D snoRNA from Cajal bodies; however, as the U2 snRNA was also absent from the Cajal bodies in these cells. The fact that in SMN depleted cells there is a reduction in the number of

Cajal bodies (Shparagel and Matera, 2005; Lemm et al, 2006) makes this data is difficult to interpret. Interestingly, the depletion of Gemin2 did not prevent the localisation of either the U3 snoRNA or U2 snRNA to the Cajal bodies. This is in agreement with previous studies that the depletion of Gemin2 does not result in a significant reduction in the number of Cajal bodies (Shparagel and Matera, 2005). This also suggests that Gemin2 is not required in processes that result in the localisation of snRNA and box C/D snoRNA to the Cajal bodies.

The data presented here further implicates the SMN complex in box C/D snoRNP biogenesis. In regards to the function of the SMN complex in box C/D snoRNP biogenesis that is still not clear; however, based on the data it is possible that the SMN complex either functions as a transport factor or in the association of fibrillarin with the box C/D snoRNP. An issue, however, with the later proposal is that unlike the siRNA mediated depletion of fibrillarin, which results in a mild reduction in box C/D snoRNA levels (Figure 3.4; Watkins et al, 2004), the depletion of the SMN protein has no effect on box C/D snoRNA levels. However, an explanation for these differences could be that if the decline in fibrillarin association caused by the loss of SMN is not significant then the levels of box C/D snoRNA may only be slightly affected. Another issue with this proposal is that siRNA mediated depletion of fibrillarin results in the accumulation of box C/D snoRNA in the nucleoplasm and Cajal bodies (Figure 3.4) while the depletion of SMN results in the accumulation of box C/D snoRNA in the nucleoplasm and cytoplasm (Figure 5.10). This indicates that SMN does not function in association of fibrillarin with the box C/D snoRNP and it is more likely that the SMN complex functions as a transport factor in box C/D snoRNP biogenesis. However, further investigation is required to determine the role of the SMN complex in box C/D snoRNP biogenesis.

The immunoprecipitation assays performed using extracts from inducible FLAG-fibrillarin HEK293 cells provide evidence as to the stoichiometry of the box C/D snoRNP complex. As it has been predicted that box C/D snoRNPs contain two copies of fibrillarin, one copy at the C/D box and another at the C'/D' box (Cahill et al, 2002; Aittaleb et al, 2003; Rashid et al, 2003; Tran et al, 2003), it was expected that endogenous fibrillarin would be co-precipitated from FLAG-fibrillarin HEK293 cell extracts using anti-FLAG antibodies. No endogenous fibrillarin was, however, co-

precipitated despite the fact that the data indicated that the FLAG-fibrillarin associates with box C/D snoRNPs (Figure 5.7). The lack of co-precipitation of endogenous fibrillarin could suggest that only one copy of fibrillarin associates with the box C/D snoRNPs. As the C'/D' box is not as conserved as the C/D box it is possible that not all box C/D snoRNPs contain two copies of fibrillarin and there may be species differences. One issue with this immunoprecipitation assay, however, is that it is possible that the FLAG-fibrillarin was expressed at a higher rate than endogenous fibrillarin. It is therefore feasible that there was an excess of FLAG-fibrillarin produced during the time frame of the experiment resulting in the formation of box C/D snoRNPs containing two copies of FLAG-fibrillarin, preventing the association of endogenous fibrillarin. This would therefore inhibit the co-precipitation of endogenous fibrillarin with FLAG-fibrillarin. Further investigations are required to determine the stoichiometry and arrangement of common core proteins on box C/D snoRNPs.

Chapter six

Analysis of the interaction of Snurportin1 with the box C/D pre-snoRNP

6.1. Introduction

The biogenesis of the box C/D snoRNPs and snRNPs share many similarities despite their differing roles in the cell. The box C/D snoRNPs function in the chemical modification of rRNA or assist in pre-rRNA processing, while the snRNPs operate in pre-mRNA splicing as part of the spliceosome. The biogenesis of the snRNPs is more defined than that of the box C/D snoRNPs and therefore provides a useful paradigm for understanding how box C/D snoRNPs are formed.

The RNA component of the box C/D snoRNPs and snRNPs (with the exception of the U6 snRNA) is transcribed by RNA polymerase II in the nucleoplasm as a precursor with 3' extended sequences. All the snRNAs are transcribed from independent genes and co transcriptionally acquire 5' m⁷G caps; however, this is not the case for box C/D snoRNAs. In vertebrates the majority of box C/D snoRNAs are encoded within the introns of house keeping genes and are released by endonucleases, resulting in intronic box C/D snoRNAs containing 5'-phosphate terminals (Tollervey and Kiss, 1997). A few vertebrate box C/D snoRNAs, such as the U3 and U8 snoRNAs, are encoded as independent genes and these have a co transcriptionally added 5' m⁷G cap (Speckmann et al, 2000; Girard et al, 2007).

The 5' m⁷G cap of both snRNAs and independently transcribed box C/D snoRNAs is converted to an m₃G cap by the hypermethylase TGS1 (Huber et al, 1998; Speckmann et al, 2000; Mitrousis et al, 2008). While the 5' m⁷G cap of snRNA is converted into an m₃G cap in the cytoplasm where this occurs with the box C/D pre-snoRNAs is not clear with both the Cajal bodies and cytoplasm implicated (Verheggen et al, 2002; Watkins et al, 2007).

The box C/D snoRNPs also associate with numerous snRNP transport factors such as CBC, CRM1, PHAX and Snurportin1; however, the function of these is best defined for the snRNPs (Yang et al, 2000; Boulon et al, 2004; Watkins et al, 2004, 2007). In snRNP biogenesis these factors are essential for the localisation of the pre-snRNA and biogenesis (Figure 6.1). For instance CBC, CRM1 PHAX interact with nuclear pre-snRNA via the m⁷G cap, which with RanGTP results in the nuclear export of the pre-

snRNA (Ohno et al, 2000; Kitao et al, 2008). In the cytoplasm the SMN complex, which has also been linked to snoRNP biogenesis (Pellizzoni et al, 2001; Jones et al 2001; see Chapter 5), facilitates the formation of the protein Sm core and recruits TGS1 resulting in the formation of the 5' m₃G cap (Ohno et al, 2000; Pellizzoni et al, 2002; Yong et al, 2002; 2004). The nuclear import protein, Snurportin1, associates with the 5' m₃G cap of the snRNA and mediates an interaction with importin β resulting in the nuclear import of the pre-snRNP complex (Huber et al, 1998; Mitrousis et al, 2008). Interestingly, SMN has also been shown to function in the nuclear import of the snRNPs in the absence of Snurportin1 (Narayanan et al, 2002. 2004). Once in the nucleus the pre-snRNP localises to the Cajal bodies where type specific snRNP core proteins are proposed to associate and the snRNPs form U4/U6 di-snRNPs and U4/U6.U5 tri-snRNPs (Stanek et al, 2003; Schaffert et al, 2004).

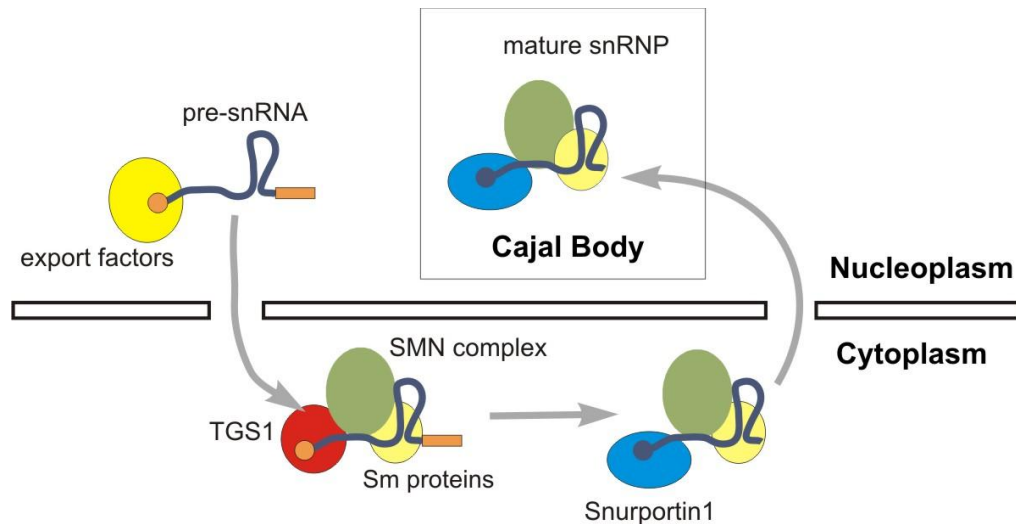


Figure 6.1: The transport of snRNPs during biogenesis

A dark grey line represents the snRNA with the 5' m⁷G cap shown as an orange circle and the 3' extended sequence as an orange rectangle. A yellow circle represents the export factors CBC, CRM1, PHAX and RanGTP. The Sm core proteins are represented by a light yellow oval. The SMN complex is represented by a green oval. TGS1 is shown as a red oval. Snurportin1 is represented as a blue oval. The 5' m₃G cap of mature snRNA is shown as by a dark grey circle. The Cajal body is represented as a grey outlined box. The nuclear membrane is shown by black outlined rectangles. Grey arrows indicate the progression of biogenesis. Figure adapted from image provided by N.J.Watkins, Newcastle University, UK.

Box C/D snoRNP biogenesis proceeds in a large dynamic complex known as the pre-snoRNP that consists of snoRNA, the common core proteins (fibrillarin, NOP56,

NOP58 and 15.5K), RNA processing and assembly factors (TIP48, TIP49, BCD1, NOP17, TAF9 and NUFIP; Watkins et al, 2004, 2007; McKeegan et al, 2007). The CBC, CRM1, PHAX and Snurportin1 factors that associate with box C/D pre-snoRNPs, most likely, function in the transport of this complex (Watkins et al, 2004; 2007, Boulon et al, 2004). However, the location(s) where box C/D snoRNP biogenesis occurs is not clear with data suggesting the nucleoplasm (Watkins et al, 2004, 2007; McKeegan et al, 2007), Cajal bodies (Boulon et al, 2007; Verheggen et al, 2002) and also the cytoplasm (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007).

Snurportin1 was first implicated in box C/D snoRNP biogenesis when it was found associated with U8 pre-snoRNPs in the cytoplasm of human cells, which contained m₃G capped pre-U8 snoRNA (Watkins et al, 2007). As Snurportin1 interacts with the m₃G cap of snRNA (Huber et al, 1998; Mitrousis et al, 2008) it is conceivable that the same interaction occurs with m₃G capped U8 snoRNA. Further evidence of the involvement of Snurportin1 in box C/D snoRNP biogenesis was shown by siRNA mediated depletion of Snurportin1 which resulted in the accumulation of U8 snoRNA outside the nucleolus in the nucleoplasm and cytoplasm (Watkins et al, 2007). The accumulation of U8 snoRNA in the cytoplasm indicates that Snurportin1 may function in the nuclear import of box C/D pre-snoRNPs.

The mechanism by which Snurportin1 associates with the cytoplasmic U8 box C/D pre-snoRNP is not known. It is possible that Snurportin1 interacts with the U8 snoRNA 5' m₃G cap; however, other interactions may also occur. To determine whether Snurportin1 interacts with other components of the box C/D pre-snoRNP a series of *in vitro* interaction assays were performed between Snurportin1 and box C/D snoRNP common core proteins and assembly factors. As only independently transcribed box C/D snoRNAs such as U3, U8 and U13 contain 5' m₃G caps by characterising the interaction of Snurportin1 with pre-snoRNP proteins it will be possible to speculate as to the involvement of Snurportin1 in the biogenesis of intronic box C/D snoRNPs. In addition, as the SMN complex has been implicated in box C/D snoRNP biogenesis (Pellizzoni et al, 2001; Jones et al, 2001) and as Snurportin1 has been shown to associate in pre-snoRNP cytoplasmic complexes, containing SMN (Narayanan et al, 2002), *in vitro* interaction assays were performed to determine whether Snurportin1 and the SMN complex interact.

6.2. Results

6.2.1. Snurportin1 interacts with the box C/D snoRNP common core protein NOP56

The box C/D pre-snoRNP is formed through an intricate network of interactions between the various components. Snurportin1 associates with cytoplasmic U8 pre-snoRNPs; however, the interaction with the pre-snoRNP has not been characterised. To determine whether Snurportin1 interacts with any box C/D snoRNP common core proteins *in vitro* interaction assays were performed.

Human Snurportin1 was expressed and purified, from *E. coli*, with an N terminal GST tag (GST-SPN1, expression plasmid provided by A.Dickmanns, Max Planck Institute of Biophysical Chemistry, Göttingen, Germany). The human box C/D snoRNP common core protein 15.5K was expressed and purified, from *E. coli*, fused to an N terminal His tag (His-15.5K, expression plasmid provided by E.S.Maxwell, North Carolina State University, USA). Due to solubility issues deletion mutants of NOP56 (1-458) and NOP58 (1-435), that lack the C terminal charged region, were expressed and purified from *E.coli* with an N terminal thioredoxin tag and C terminal His tag (His-NOP56, His-NOP58; McKeegan et al, 2007). The C terminal charged region of NOP56 and NOP58, which was removed in the deletion mutants, has been shown not to be required for box C/D snoRNP formation in yeast (Lafontaine and Tollervey, 2000). Unfortunately, it was also not possible to express and purify full length fibrillarin; however, it was possible to express and purify a deletion mutant, which lacked the N terminal RGG motif, fused to an N terminal His tag (His-N-C-fibrillarin (amino acids 81-321); see Chapter 4.2.1).

GST-Snurportin or GST alone (negative control) were immobilised on glutathione Sepharose beads and incubated with His tagged proteins. After two hours incubation the glutathione Sepharose beads were washed and bound proteins analysed by separation on a 12 % SDS-PAGE gel followed either by Coomassie blue staining of the gel (Figure 6.2 A) or Western blot assays using anti-His antibodies (Figure 6.2 B).

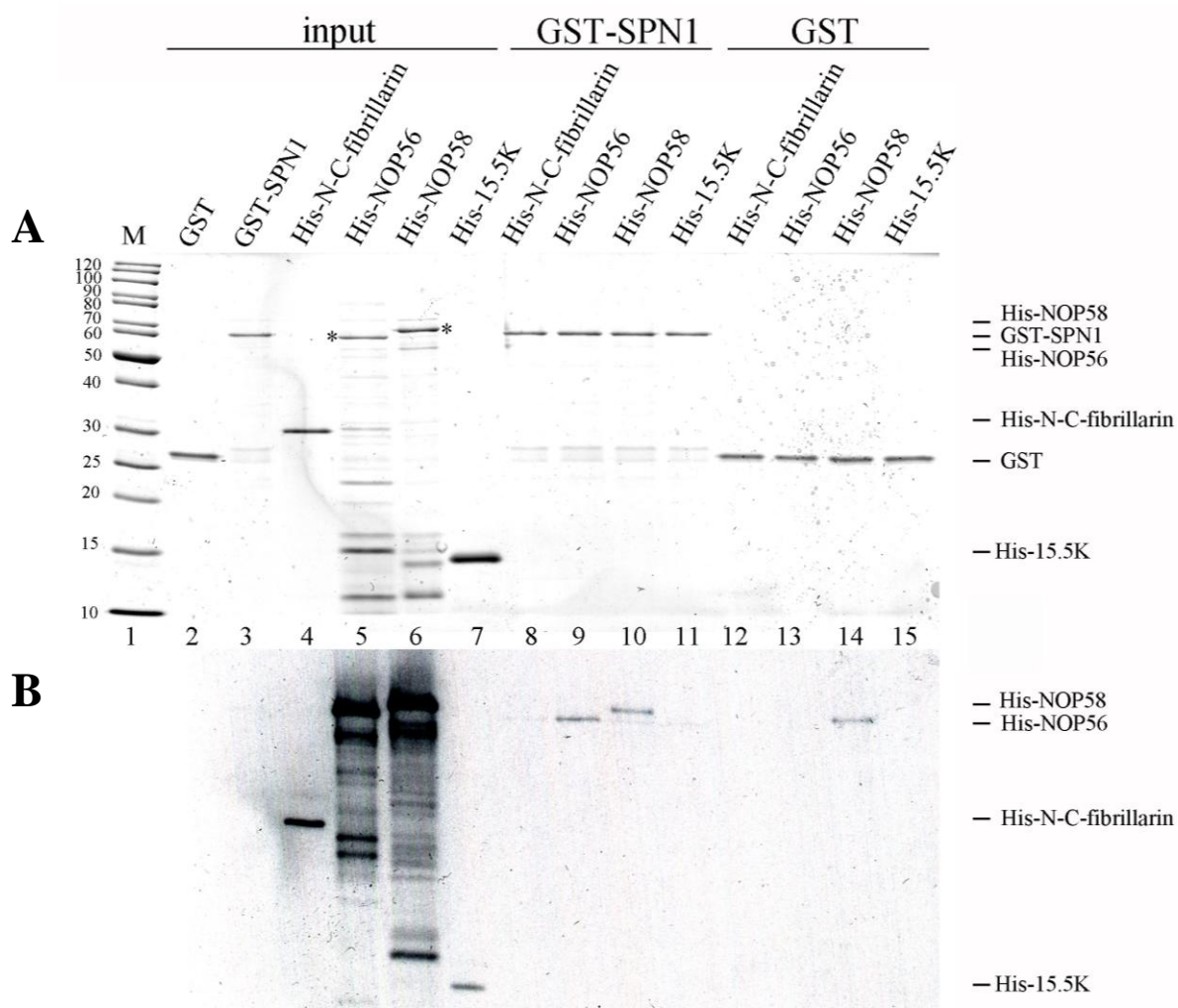


Figure 6.2: Interaction assay between Snurportin1 and the box C/D snoRNP common core proteins

GST-Snurportin1 (GST-SPN1) or GST alone was immobilised on glutathione Sepharose beads and incubated with His-N-C-fibrillarin, His-NOP56, His-NOP58 and His-15.5K. Bound proteins were separated on a 12 % SDS-PAGE followed by either (A) Coomassie blue staining or (B) Western blot analysis using anti-His antibodies. Where appropriate full length proteins are indicated by an asterisk. Inputs represent 10 % of protein used in the interaction assay. The molecular markers (M) sizes are shown to the left of the panel in kilodaltons. The identity of the GST and His tagged proteins used is indicated at the top of each lane. The migration of the individual proteins is indicated to the right of the panel.

Immobilised GST-Snurportin1 only retained the His-NOP56 above the levels of the GST control (Figure 6.2 B). There was retention of His-NOP58 by GST-Snurportin1; however, this was not above the background levels seen with the GST control. This indicates that out of the common core proteins Snurportin1 specifically interacts with NOP56. As the interaction of Snurportin1 and NOP56 was only detectable using the more sensitive Western blot assay (Figure 6.1 B) this suggests that the interaction is either weak and / or transient.

6.2.2. Snurportin1 interacts with the box C/D snoRNP assembly factors TIP48 and TIP49

To further analyse the interactions of Snurportin1 with the box C/D pre-snoRNP *in vitro* interaction assays were performed between Snurportin1 and the box C/D snoRNP assembly factors TIP48 and TIP49. GST-Snurportin and GST alone were immobilised on glutathione Sepharose beads and incubated with recombinant human TIP48 and TIP49, which were expressed and purified with an N terminal His tag from *E. coli* (His-TIP48 and His-TIP49; McKeegan et al, 2007). In addition, as TIP48 and TIP49 are ATPases, the interaction assays were also performed in the presence of 100 mM ATP or ADP to determine whether nucleotides were required for interactions.

Immobilised GST-Snurportin1 retained His-TIP48 above the levels seen with GST alone, both in the presence and absence of ATP and ADP (Figure 6.3). Interestingly, the presence of ATP or ADP reduced the retention of His-TIP48 by both GST-Snurportin1 and the GST control. As the retention of His-TIP48 by GST-Snurportin1 was highest in the absence of ATP or ADP this indicates that the presence of nucleotides is not required for this interaction. Since the retention of His-TIP48 by GST-Snurportin1 was visible by both Coomassie blue staining (Figure 6.3 A) as well as the more sensitive Western blot assay (Figure 6.3 B) this suggests that the interactions were strong.

His-TIP49 was not retained by the GST negative control; however, was retained by GST-Snurportin1 (Figure 6.3 A and B). Interestingly, there was a reduction in the retention of His-TIP49 by GST-Snurportin1 in the presence of ATP or ADP. This indicates that the interaction between His-TIP49 and GST-Snurportin1 does not require nucleotide. As retention of His-TIP49 by immobilised GST-Snurportin1 was visible using Coomassie blue staining (Figure 6.3 A) as well as the more sensitive Western blot assay (Figure 6.3 B) this indicates that the interaction between GST-Snurportin1 and His-TIP49 was a strong stable interaction.

Taken together these results indicate that Snurportin1 interacts with both TIP48 and TIP49 and these interactions do not require nucleotide. The fact that these interactions were seen by direct Coomassie blue staining of the gel as well as the more sensitive Western blot assay indicate that these are strong stable interactions.

6.2.3. Snurportin1 does not interact with the box C/D snoRNP assembly factors NUFIP, TAF9, NOP17 or the transport factor PHAX

Further *in vitro* interaction assays were performed to determine whether Snurportin1 interacts with any other box C/D snoRNP assembly factors or the transport factor PHAX. Unfortunately, it was not possible to express and purify these assembly factors fused to N terminal His tags. Therefore the assembly factors NUFIP, TAF9 and NOP17 and the transport factor PHAX were expressed using reticulocyte lysate in the presence of [³⁵S] methionine (McKeegan et al, 2007). The interaction assays were performed as described earlier with the exception that results were visualised by autoradiography once proteins were separated on a 12 % SDS-PAGE gel.

Immobilised GST-Snurportin1 failed to retain the *in vitro*-translated ³⁵S-labelled NUFIP, TAF9, NOP17, BCD1 and PHAX proteins above the levels of GST alone, which indicates that Snurportin1 does not interact with any of these factors (Figure 6.4).

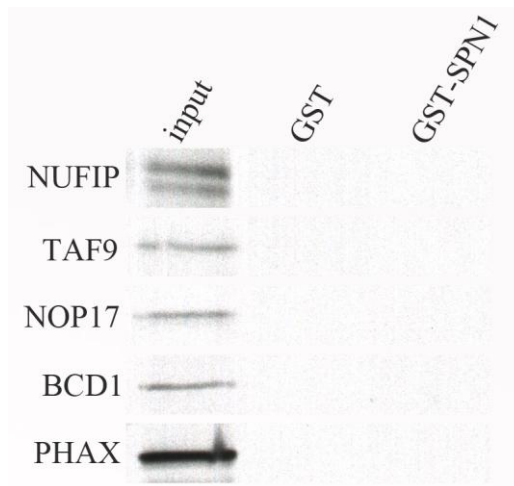


Figure 6.4: Interaction assay between Snurportin1 and the box C/D snoRNP assembly factors and the transport factor PHAX

GST-Snurportin1 (GST-SPN1) or GST alone was immobilised on glutathione Sepharose beads and incubated with *in vitro*-translated ^{35}S -labelled NUFIP, TAF9, NOP17, BCD1 or PHAX. Bound proteins were separated on a 12 % SDS-PAGE gel and results visualised by autoradiography. Input lane represents 10 % of *in vitro*-translated ^{35}S -labelled protein used in interaction assay. The identity of GST tagged protein used is indicated above each lane. The identity of the radiolabelled protein used is indicated to the left of the panel.

6.2.4. Snurportin1 does not interact with the SMN complex

As the SMN complex and Snurportin1 have been implicated in box C/D snoRNP biogenesis (Pellizzoni et al, 2001; Jones et al, 2001; Watkins et al, 2007) and are both found in pre-snRNP import complexes (Narayanan et al, 2002) *in vitro* interaction assays were performed to determine whether Snurportin1 interacts with any of the SMN complex proteins.

The interaction assays were performed as described earlier with immobilised GST-Snurportin1 or GST alone incubated with SMN complex proteins that were expressed using reticulocyte lysate in the presence of [^{35}S] methionine (expression plasmids provided by U.Fischer, University of Wurzburg, Germany).

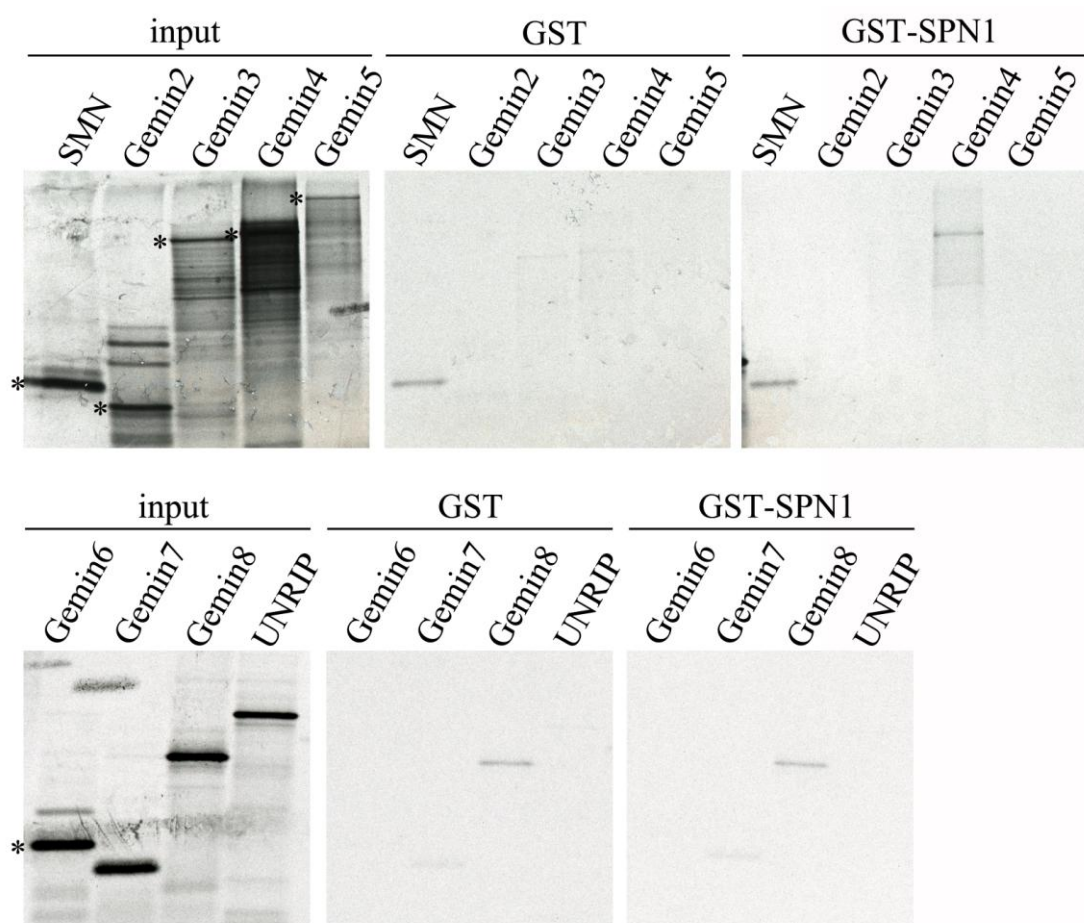


Figure 6.5: Interaction assay between Snurportin1 and components of the SMN complex

GST-Snurportin1 (GST-SPN1) or GST alone was immobilised on glutathione Sepharose beads and incubated with *in vitro*-translated ^{35}S -labelled SMN, Gemin2-8 or UNRIP. Bound proteins were separated on a 12 % SDS-PAGE gel and results visualised by autoradiography. Input lane represents 10 % of *in vitro*-translated ^{35}S -labelled protein used in interaction assay. The identity of the GST tagged protein and radiolabelled protein used is indicated above each lane. Asterisk indicates the full length protein where necessary.

Immobilised GST-Snurportin1 did not retain *in vitro*-translated ^{35}S -labelled SMN, Gemin2, 3, 5, 6, 7, 8 or UNRIP above the GST control level, which indicates that no interaction occurs between these factors (Figure 6.5). The only SMN complex factor that immobilised GST-Snurportin1 retained above GST levels was Gemin4. The retention of Gemin4 by GST-Snurportin1 was low and not convincing and therefore further investigation is required to determine whether the SMN complex interacts with Snurportin1.

6.3. Discussion

Snurportin1 was first implicated in box C/D snoRNP biogenesis when it was found associated with cytoplasmic U8 pre-snoRNP complexes containing m₃G capped snoRNA (Watkins et al, 2007). As Snurportin1 has previously been shown to bind the m₃G caps of snRNAs it is conceivable that Snurportin1 associates with cytoplasmic U8 pre-snoRNPs through an interaction with the 5' m₃G cap of the snoRNA. However, as intronic box C/D snoRNAs do not possess 5' m₃G caps this would indicate that Snurportin1 only functions in the biogenesis of independently transcribed snoRNAs.

To further characterise the interactions of Snurportin1 with the box C/D pre-snoRNP a series of *in vitro* interaction assays were performed between Snurportin1 and box C/D snoRNP core proteins, assembly factors and the transport factor PHAX. The *in vitro* interaction assays indicate that Snurportin1 interacts with the common core protein NOP56 (Figure 6.2) and with the box C/D snoRNP assembly factors TIP48 and TIP49 (Figure 6.3). As TIP48 and TIP49 are known ATPases the requirement of ATP and ADP for the interaction with Snurportin1 was also investigated and revealed that the presence of nucleotide is not required. The interactions of Snurportin1 with NOP56, TIP48 and TIP49 may represent the means, in addition to the 5' m₃G cap, by which Snurportin1 associates with U8 pre-snoRNP complexes. Since intronic snoRNAs, such as U14, do not possess a 5' m₃G cap the identification of these interactions show a mechanism by which Snurportin1 could associate with box C/D pre-snoRNPs that lack snoRNA with a 5' m₃G cap and thus function in their biogenesis. Further investigation is, however, required to determine if Snurportin1 associates with other species of box C/D pre-snoRNP.

As SMN and Snurportin1 have been implicated in box C/D snoRNP biogenesis (Pellizzoni et al, 2001; Jones et al, 2001) and are both found in pre-snoRNP import complexes (Narayanan et al, 2002) *in vitro* interaction assays were performed to determine whether Snurportin1 interacts with the SMN complex. These assays indicate that Snurportin1 and the SMN complex do not directly interact. There was a slight retention of Gemin4 by GST-Snurportin1 in the *in vitro* interaction assays; however, this was weak and far from conclusive (Figure 6.5). As other studies have indicated that

Snurportin1 and the SMN complex do not directly interact but are in the same snRNP import complex through a mutual association with the snRNA this suggests that these factors, more than likely, do not interact (Narayanan et al, 2002).

The most likely function of Snurportin1 in box C/D snoRNP biogenesis is in the transport of cytoplasmic pre-snoRNPs to the nucleus. This is indicated by the fact that Snurportin1 was found specifically associated with cytoplasmic U8 pre-snoRNPs (Watkins et al, 2007), functions in the nuclear import of snRNPs (Huber et al, 1998) and the siRNA mediated depletion of Snurportin1 results in the accumulation of U8 snoRNA in the cytoplasm (Watkins et al, 2007). It is also possible that Snurportin1 could function in the transport of the pre-snoRNP to the Cajal bodies as Snurportin1 has been shown to localise to these sites during snRNP biogenesis (Ospina et al, 2005).

Chapter seven

General discussion

7. Discussion

7.1. Overview

This study set out to further characterise box C/D snoRNP biogenesis. This included analysis of the interactions of the fibrillarin common core protein with the box C/D snoRNP assembly factors and the importance of the various domains of fibrillarin for localisation and association with the snoRNPs. Furthermore, as fibrillarin has been shown to interact with the SMN protein (Pellizzoni et al, 2001; Jones et al, 2001) the role of the SMN complex in box C/D snoRNP biogenesis was also investigated.

7.2. Box C/D snoRNP biogenesis

The association of the common core proteins to the box C/D snoRNP is a hierarchical process with the binding of 15.5K to the snoRNA essential for the subsequent association of fibrillarin, NOP56 and NOP58 (Watkins et al, 2002). Studies in yeast have shown that Snu13p (15.5K) and Nop58p (NOP58) are vital for the accumulation of box C/D snoRNA (Lafontaine and Tollervey, 1999; Watkins et al, 2000). In this study it was also shown that of the human common core proteins only the loss of NOP58 results in a dramatic reduction in box C/D snoRNA levels; however, it was not possible to test 15.5K. Taken together as 15.5K and NOP58 are essential for the accumulation of box C/D snoRNA this could indicate that these common core proteins form a basic complex with the snoRNA enhancing its stability, perhaps by preventing degradation from exonucleases (Lafontaine and Tollervey, 1999; Watkins et al, 2000, 2004).

Numerous factors have been identified in eukaryotes that are required for the assembly of the box C/D snoRNPs (Peng et al, 2003; Wang and Meier, 2004; Boulon et al, 2004, 2008; Watkins et al, 2004, 2007; Gonzales et al, 2005; McKeegan et al, 2007). In this study multiple fibrillarin deletion mutants were used to analyse the interactions of fibrillarin with the box C/D snoRNP assembly factors. Protein-protein interaction studies revealed that the C terminal domain of fibrillarin (amino acids 141-321) mediates interactions with the assembly factors BCD1, NOP17, NUFIP, TIP48 and TIP49. Previous studies have shown that human fibrillarin also interacts with NOP56

(McKeegan et al, 2007) and investigations using archaeal proteins revealed that the fibrillarin C terminal domain interacts with the NOP56 / NOP58 homologue, Nop5 (Aittaleb et al, 2003). It is therefore possible that the interaction of eukaryotic fibrillarin with NOP56 is also mediated by the fibrillarin C terminal domain. In this study the further use of the fibrillarin deletion mutants showed that the C terminal domain of fibrillarin was also essential for the correct cellular localisation, which is in agreement with previous studies (Snaar et al, 2000). As the fibrillarin C terminal domain is important for mediating interactions with the assembly factors (and potentially NOP56) and vital for localisation it is possible that this domain is required for the association of fibrillarin with the box C/D snoRNP (see future work).

The site(s) of box C/D snoRNP biogenesis are not clear with data implicating the nucleoplasm (Terns et al, 1995; Watkins et al, 2004; Boulon et al, 2004), Cajal bodies (Verheggen et al, 2001; Boulon et al, 2004; Lemm et al, 2006) and cytoplasm (Baserga et al, 1992; Peculis et al, 2001; Watkins et al, 2007). In this study the localisation of the box C/D snoRNP assembly factors was investigated to provide further data on the site(s) of box C/D snoRNP biogenesis. Analysis in human cells revealed that the box C/D snoRNP assembly factors BCD1, NUFIP, TAF9, TIP48, and TIP49 were distributed throughout the nucleoplasm and not enriched in any nuclear bodies. This strengthens the case that box C/D snoRNP biogenesis occurs in the nucleoplasm and is in agreement with previous studies that have shown these assembly factors associate with nucleoplasmic box C/D pre-snoRNPs (Watkins et al, 2004, 2007; McKeegan et al, 2007). The nucleoplasm and Cajal bodies were further implicated as sites of box C/D snoRNP biogenesis in this study as the depletion of NOP56 and / or fibrillarin in human cells resulted in the accumulation of the U3 box C/D snoRNA in these locations (McKeegan et al, 2007). The accumulation of U3 snoRNA in the nucleoplasm and Cajal bodies could indicate that only fully assembled box C/D snoRNPs can leave these sites or alternatively it is possible that these are the locations where box C/D snoRNPs accumulate when assembly cannot be completed.

The Cajal bodies have been shown to be the location of U4/U6 di-snRNP and U4/U6.U5 tri-snRNP formation (Stanek et al, 2003; Schaffert et al, 2004) and have also been proposed as the site of the association of the snRNP core proteins (Nesic et al, 2004). The Cajal bodies have been implicated in box C/D snoRNP biogenesis in a

number of studies (Verheggen et al, 2001; Boulon et al, 2004; Lemm et al, 2006); however, the purpose of the localisation of box C/D snoRNA to these sites is not clear (Naryanan et al, 1999). It has been proposed that the Cajal bodies are site where the box C/D snoRNP common core proteins are associated with the snoRNA and also the location where 5' m⁷G cap of U3 box C/D snoRNA is converted to an m₃G cap; however, further investigation is required to determine this (Verheggen et al, 2002; Boulon et al, 2004). In this study the observation that the U3 box C/D snoRNA accumulates in the Cajal bodies of NOP56 and fibrillarin depleted human cells provides further evidence that the Cajal bodies are sites of box C/D snoRNP biogenesis. Additional study is required to determine the purpose of box C/D snoRNA localisation to the Cajal bodies. Whether box H/ACA snoRNP biogenesis also occurs in the Cajal bodies is not clear as it has yet to be demonstrated that box H/ACA snoRNAs transiently localise to these nuclear bodies (Kiss, 2004). Furthermore, while the box H/ACA snoRNP core proteins such as GAR1 are found in Cajal bodies these may represent scaRNPs (Jady and Kiss, 2001).

It has been proposed that only fully formed mature box C/D snoRNPs localise to the nucleolus as the depletion of yeast Snu13p, Nop56p, Nop58p and Nop1p inhibit the nucleolar localisation of box C/D snoRNAs (Verheggen et al, 2001). As depletion of Nop58p results in the dramatic reduction of box C/D snoRNA levels artificial snoRNAs with long terminal stems were used to determine that it was required for the nucleolar localisation of the box C/D snoRNPs (Verheggen et al, 2001). In this study depletion of NOP56 and fibrillarin (15.5K not tested) in human cells resulted in an accumulation of the U3 box C/D snoRNA in the nucleoplasm and Cajal bodies. Furthermore the depletion of NOP56 resulted in the accumulation of newly synthesised fibrillarin (expressed from a reporter plasmid) in the nucleoplasm. The depletion of NOP58 resulted in such a dramatic reduction in box C/D snoRNA levels that it is not possible to determine whether it was required for nucleolar localisation. Nevertheless, in combination with the yeast data (Verheggen et al, 2001) these results indicate that a full set of common core proteins is required for the nucleolar localisation of box C/D snoRNPs. Whether the assembled mature box C/D snoRNP is itself a nucleolar localisation signal or that associated transport factors transfer the box C/D snoRNPs to the nucleolus is not clear. However, it has been demonstrated that the transport factor CRM1 is required for the nucleolar localisation of U3 box C/D snoRNA (Boulon et al,

2004) and Nopp140, which has been linked to box C/D snoRNP biogenesis (Yang et al, 2000), has been shown to shuttle between the nucleolus and cytoplasm in mammalian cells (Meier and Blobel, 1992). Further investigation is required to determine the mechanism of nucleolar localisation.

7.3. The SMN complex functions in the biogenesis of numerous RNPs

One of the main aims of this study was to determine the role of the SMN complex in box C/D snoRNP biogenesis. The SMN complex has been linked to the biogenesis of numerous RNPs, including the snoRNPs (Pellizzoni et al, 2001; Jones et al, 2001), telomerase RNPs (Bachand et al, 2002) and miRNPs (Mourelatos et al, 2002); however, its function has been best defined in snRNP biogenesis (Yong et al, 2002, 2004; Chari et al, 2008).

One of the functions of the SMN complex in snRNP biogenesis is facilitating the binding of the Sm protein core to the snRNAs (Chari et al, 2008). The SMN complex was first implicated in box C/D snoRNP biogenesis through the identification of an interaction between the SMN protein and fibrillarin, which was also shown in this study (Pellizzoni et al, 2001; Jones et al, 2001). Interestingly, in this study the SMN protein was also found to interact with the box C/D snoRNP assembly factors BCD1 and NUFIP, suggesting that SMN associates with the box C/D pre-snoRNP. Based on the interactions of SMN with fibrillarin and the pre-snoRNP it could be postulated that the SMN complex functions in the recruitment of fibrillarin to the box C/D pre-snoRNP. As fibrillarin and box C/D snoRNA are found enriched in Cajal bodies (Tollervey et al, 1993; Narayanan et al, 1999) and SMN is found in both Cajal bodies and in Gems (which are sometimes associated with Cajal bodies) (Liu and Dreyfuss, 1996) it is tempting to propose that SMN functions in fibrillarin recruitment to the pre-snoRNP in these nuclear bodies. The assembly factors that SMN interacts with are, however, not enriched in Cajal bodies but found throughout the nucleoplasm, which indicates that they are associated only with nucleoplasmic pre-snoRNPs. Although the presence of the assembly factors in the Cajal bodies cannot be excluded it does indicate that if the SMN complex functions in the recruitment of fibrillarin to the box C/D pre-snoRNP then it does not occur in the Cajal bodies.

There are a number of issues with the proposal that SMN functions in the recruitment of fibrillarin with the box C/D pre-snoRNP or in other processes required for the association of fibrillarin with the box C/D snoRNA. For instance the siRNA mediated depletion of fibrillarin and SMN do not affect box C/D snoRNA localisation and levels in the same manner. This could, however, reflect a difference between the loss of fibrillarin and the lack of association with the box C/D snoRNP. Another issue with this proposal is that the deletion of the SMN interacting domain of fibrillarin (amino acids 1-81) does not affect the localisation of fibrillarin. Further work is, however, required to determine whether in the absence of the SMN interacting domain fibrillarin can associate with the box C/D snoRNP (see future work).

The SMN complex has also been proposed to function as a nuclear import factor (Narayanan et al, 2002, 2004). The snRNP nuclear import complex consists of the pre-snRNP, SMN complex, Snurportin1 and importin β , and interestingly it has been shown that SMN can function in the nuclear import of pre-snRNPs independently of Snurportin1 (Narayanan et al, 2002, 2004). As some box C/D pre-snoRNPs, such as U8, U3 and U13 (Watkins et al, 2007), are found in the cytoplasm the SMN complex could function in their nuclear import (Figure 7.1). Also as the SMN complex localises to the Cajal bodies it could also function in the transport of box C/D pre-snoRNPs to this nuclear body (Figure 7.1). This proposal is supported by siRNA mediated depletion of components of the SMN complex, which results in the accumulation of box C/D snoRNA in the nucleoplasm and in some cases the cytoplasm.

The SMN complex and Snurportin1 have been implicated in box C/D snoRNP biogenesis (Pellizzoni et al, 2001; Jones et al, 2001; Watkins et al, 2007) and are both found in pre-snRNP import complexes (Narayanan et al, 2004). In this study it was revealed that the SMN complex and Snurportin1 do not interact directly. This is in agreement with previous studies that have shown that the SMN complex and Snurportin1 are both associated with the pre-snRNP import complex through interactions with the snRNA (Narayanan et al, 2004). Interestingly, in this study Snurportin1 was shown to interact with the box C/D snoRNP common core protein NOP56 and the assembly factors TIP48 and TIP49, which may represent, in addition to the m₃G cap of independently transcribed box C/D snoRNAs, how Snurportin1 interacts

with box C/D pre-snoRNPs. Further investigation is, however, required to determine if Snurportin1 interacts with other box C/D pre-snoRNPs and whether the m₃G cap is required.

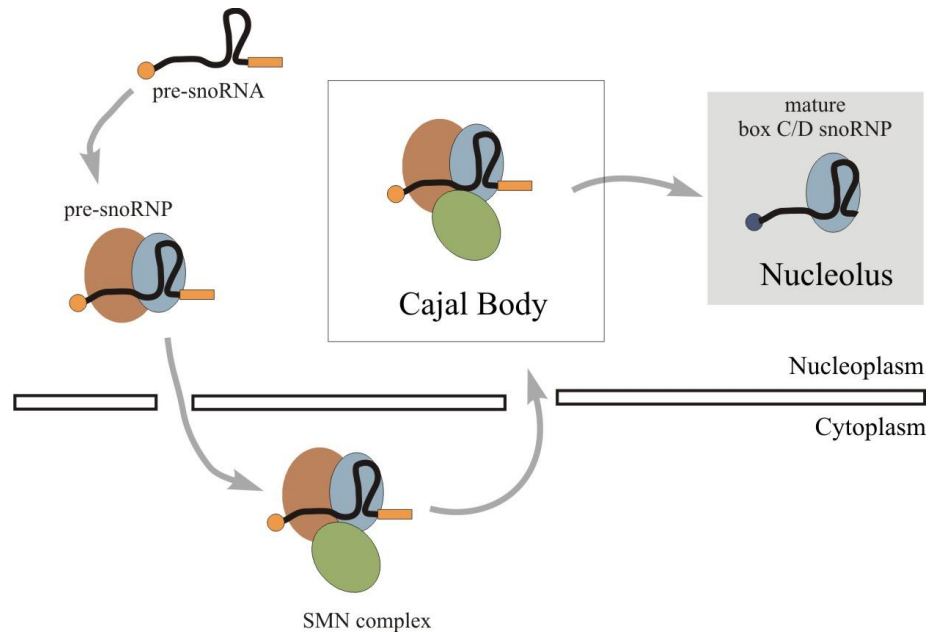


Figure 7.1: SMN could function as a transport factor in box C/D snoRNP biogenesis

The box C/D snoRNA is represented as a black line with the 5' m⁷G cap shown as an orange circle and the 3' extended sequence as an orange box. The box C/D snoRNP core proteins are shown as a blue oval while the assembly, RNA processing and transport factors are represented by a brown oval. The SMN complex is represented as a green oval. The 5' m₃G cap of mature snoRNA is shown as a blue circle. Grey arrows show the progression of biogenesis. The nuclear membrane is represented by black outlined rectangles. Figure adapted from image provided by N.J.Watkins, Newcastle University, UK.

This study strengthens the case that the SMN complex is involved in box C/D snoRNP biogenesis despite not being able to clearly identify its role. Based on the data presented the SMN complex, more than likely, functions in the transport of the box C/D snoRNPs rather than in the association of fibrillarin with the snoRNP (Figure 7.1). It is, however, necessary to establish *de novo* biogenesis assays before it can be concluded that the SMN complex is not involved in box C/D snoRNP protein assembly. The use of *de novo* biogenesis assays was essential for determining the role of the SMN complex in snRNP biogenesis (Shparagel and Matera, 2005) and therefore a variety of *de novo*

biogenesis assays were developed in this study; however, for a variety of reasons these were unsuccessful (see future work).

Further analysis of the SMN complex is required, as the mutation of the gene encoding the SMN protein has been linked to the neurodegenerative disease SMA. While the genetics of this disease are well defined the molecular mechanisms that result in the specific degeneration of the motor neurons are not. A candidate process for SMA is the disruption of pre-mRNA splicing due to issues with snRNP biogenesis (Winkler et al 2005; Shpargel and Matera, 2005; Zhang et al, 2008); however, the SMN complex has also been linked to the biogenesis of the box C/D and H/ACA snoRNPs (Pellizzoni et al, 2000; Jones et al, 2001), telomerase RNPs (Bachand et al, 2002) and miRNPs (Mourelatos et al, 2002). This indicates that the SMN complex functions in the biogenesis of a range of small RNPs and therefore further investigation is required to determine whether the dysfunction of these processes also contribute to SMA pathology.

7.4. Implications of study

The characterisation of box C/D snoRNP biogenesis provides information that can be used as a basis for analysing other RNPs. For instance NUFIP and TIP48 have been implicated in box H/ACA snoRNP biogenesis; however, their function is not as well characterised as in box C/D snoRNP biogenesis (King et al, 2001; Boulon et al, 2008). Analysis of box C/D snoRNP biogenesis has shown that NUFIP and TIP48 associate with nucleoplasmic pre-snoRNPs and form an intricate network of interactions with the common core proteins (Watkins et al, 2004, 2007; McKeegan et al, 2007). Furthermore, NUFIP has been shown to mediate the assembly of a partial box C/D snoRNP consisting of NOP56, 15.5K and snoRNA (McKeegan et al, 2007). It has therefore been proposed that NUFIP and TIP48 function in the recruitment and association of the common core proteins with the box C/D snoRNA (McKeegan et al, 2007). It is possible that NUFIP and TIP48 have similar functions in box H/ACA snoRNP biogenesis.

The biogenesis of the box C/D snoRNPs and snRNPs share many similarities, for instance both utilise the same RNA processing (exosome), RNA modification (TGS1)

and transport factors (CRM1, PHAX, Snurportin1; Huber et al, 1998; Allmang et al, 1999; Ohno et al, 2000; Verheggen et al, 2002; Boulon et al, 2004; Watkins et al, 2004, 2007). The identification that the SMN complex functions in box C/D snoRNP biogenesis, in this study and others (Pellizzoni et al, 2001; Jones et al, 2001; Lemm et al, 2006), draws another parallel between the biogenesis of the box C/D snoRNPs and snRNPs. Furthermore, as the SMN protein also interacts with the box H/ACA snoRNP core protein GAR1 (Pellizzoni et al, 2001; Jones et al, 2001) it is possible that the SMN complex has a similar function in the biogenesis of both the box C/D and box H/ACA snoRNPs.

The biogenesis of the box C/D snoRNPs, which only consist of five components (snoRNA and four common core proteins) and yet is an extremely complex process requiring numerous assembly, RNA processing and transport factors (for localisation to numerous cellular domains), serves as a paradigm for the biogenesis of larger RNPs such as the spliceosome and ribosome. For instance the box C/D snoRNP common core protein 15.5K is also a component of the U4 snRNP and is required for the formation of the U4/U6.U5 tri-snRNPs, which are the central components of the spliceosome. The incorporation of common core proteins in box C/D snoRNP biogenesis is a hierarchical process with the association of 15.5K to the C/D box of the snoRNA required for the subsequent binding of fibrillarin, NOP56 and NOP58 (Watkins et al, 2002). It has been proposed that the association of 15.5K with the snoRNA causes alterations in the RNA structure presenting additional binding sites (Watkins et al, 2002). The formation of the U4/U6.U5 tri-snRNP is also a hierarchical process, which is initiated by the association of 15.5K with the 5' stem loop of the U4 snRNA, which results in the subsequent binding of hPRP31 and the formation of the U4/U6 di-snRNP (Nottrott et al, 1999, 2002; Schaffert et al, 2004). The U5 snRNP is then incorporated to form the tri-snRNP through an interaction between the U5 snRNP hPRP6 (102K) and the U4 snRNP hPRP31 protein (Makarova et al, 2002). Therefore in both the formation of the spliceosome and box C/D snoRNPs 15.5K has a crucial role in their assembly, most likely through structural rearrangements of RNA.

The analysis of the biogenesis of small RNPs also provides concepts and mechanisms that can be extrapolated to analyse the synthesis of the ribosome. For instance the human ribosomal proteins L7A (RPL7A) and L30 (RPL30) are members of the L7Ae

protein family, which includes the box C/D and H/ACA snoRNP core proteins 15.5K and NHP2, respectively (Koonin et al, 1994; Watkins et al, 2000). Therefore analysis of the association of 15.5K with the box C/D snoRNP could provide data on the mechanisms of ribosomal protein association.

RNPs have been implicated in a variety of diseases, for instance the U50 box C/D snoRNP is a candidate tumour suppressor and mutation of the snoRNA sequence has been linked to prostate cancer (Dong et al, 2008). The box H/ACA snoRNPs and telomerase RNPs have been linked to the bone marrow failure syndrome dyskeratosis congenital (Heiss et al, 1998). In addition the telomerase RNPs have been implicated in cancer, aging and cardiovascular disease (Oulton and Harrington, 2000). The snRNPs, which make up the spliceosome, have been linked to SMA (Eggert et al, 2006; Zhang et al, 2008) and retinitis pigmentosa (Vithana et al, 2001). Furthermore the biogenesis and function of the ribosome has also been implicated in diseases such as diamond-blackfan anemia (Flygare and Karlsson, 2006), Treacher-Collins syndrome (Trainor et al, 2009) and a variety of cancers (Dai and Lu, 2008) and is also a target of a number of antibiotics. Therefore the further investigation into the biogenesis and the function of RNPs is required to understand the molecular mechanisms of numerous human diseases.

7.5. Future studies

In this study the C terminal domain of fibrillarin was shown to be essential for the localisation of fibrillarin and for interactions with the box C/D snoRNP assembly factors. It was, however, not possible to determine whether the C terminal domain was required for the association of fibrillarin with the box C/D snoRNP due to issues with the transient transfection of the GFP tagged fibrillarin deletion mutants. An inducible HEK293 cell line that expresses FLAG tagged fibrillarin was developed in this study and numerous control experiments showed that FLAG-fibrillarin was expressed at levels comparable to endogenous fibrillarin, localised correctly and associated with box C/D snoRNPs (Chapter 3.2.6). Therefore to determine whether the C terminal domain of fibrillarin is essential for association with the box C/D snoRNPs inducible HEK293 cell lines are being developed (by the lab) that express FLAG tagged fibrillarin deletion mutants. These inducible cell lines can also be used to determine whether the SMN

interacting domain of fibrillarin (amino acids 1-81) is required for association with the box C/D snoRNPs, providing further data on the role of SMN in box C/D snoRNP biogenesis.

Despite strengthening the case that the SMN complex functions in box C/D snoRNP biogenesis further work is required to clarify its role. One issue that hindered the analysis of box C/D snoRNP biogenesis and the role of the SMN complex was the inability to produce *de novo* biogenesis assays. Inducible U3 and U14 box C/D snoRNA HEK293 cell lines were developed in this study; however, it was not possible to express either box C/D snoRNA at a high enough level to be comfortably detected (Chapter 3.2.5). It may be possible to overcome the issue of low snoRNA expression by the incorporation of tandem repeats of the snoRNA sequence into these cell lines. The use of tandem repeats of snoRNA genes was used by Darzacq et al 2006 to produce an inducible cell line that expresses the E3 box H/ACA snoRNA at levels similar to endogenous E3.

Inducible FLAG-SMN and FLAG-fibrillarin HEK293 cell lines were produced in this study and used to analyse the *in vivo* interactions of SMN with fibrillarin and the pre-snoRNP. These cell lines were also developed to be used in experiments that utilised siRNA mediated depletion; however, it was not possible to efficiently deplete target proteins using siRNAs in these cell lines. The inability to specifically deplete proteins prevented the further use of these cell lines to analyse box C/D snoRNP biogenesis (Chapter 3.2.6). The targeted depletion of proteins in HEK293 cells has been performed in numerous studies (Lakkaraju et al, 2008); therefore, other methods of protein depletion are being investigated (by the lab) including the use of shRNA constructs.

In this study Gemin5 was shown to be essential for the accumulation of box C/D snoRNA, which suggests that it is involved in box C/D snoRNP biogenesis. Gemin5 does not interact with any of the box C/D snoRNP core proteins or assembly factors; however, Gemin5 has been shown to bind snRNAs (Battle et al 2006, 2007). Therefore, it is possible that Gemin5 interacts directly with box C/D snoRNA. Further investigations are therefore necessary to determine whether Gemin5 functions in box C/D snoRNP biogenesis by directly associating with the snoRNA.

The SMN complex is found in Cajal bodies (Liu and Dreyfuss, 1996), which are enriched in the scaRNPs (Jady and Kiss, 2001). As the SMN protein interacts with fibrillarin (Pellizzoni et al, 2001; Jones et al, 2001), which is a core protein of box C/D scaRNPs (Jady and Kiss, 2001), it may function in their biogenesis. Furthermore, as the scaRNPs function in the chemical modification of snRNA and the SMN complex associates with snRNPs it is possible that the SMN complex recruits scaRNPs to the snRNA. Further investigation is therefore required to determine whether the SMN complex is involved in the biogenesis and / or function of the scaRNPs.

As SMA is caused by mutation of the SMN1 gene it would be interesting to determine whether in primary fibroblast cells derived from SMA patients' box C/D snoRNP and ribosome biogenesis are affected. If the biogenesis of the box C/D snoRNPs and ribosome were affected in these cells it would provide another potential disease mechanism.

References

- Ach, R. A. & Weiner, A. M. (1991). Cooperation between CCAAT and octamer motifs in the distal sequence element of the rat U3 small nucleolar RNA promoter. *Nucleic Acids Res* **19**, 4209-18.
- Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. & Luhrmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation *in vitro*. *EMBO J* **18**, 5789-802.
- Aittaleb, M., Rashid, R., Chen, Q., Palmer, J. R., Daniels, C. J. & Li, H. (2003). Structure and function of archaeal box C/D sRNP core proteins. *Nat Struct Biol* **10**, 256-63.
- Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. & Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J* **18**, 5399-410.
- Allmang, C., Mitchell, P., Petfalski, E. & Tollervey, D. (2000). Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res* **28**, 1684-91.
- Andersen, J. S., Lam, Y. W., Leung, A. K., Ong, S. E., Lyon, C. E., Lamond, A. I. & Mann, M. (2005). Nucleolar proteome dynamics. *Nature* **433**, 77-83.
- Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., Lam, Y. W., Steen, H., Mann, M. & Lamond, A. I. (2002). Directed proteomic analysis of the human nucleolus. *Curr Biol* **12**, 1-11.
- Anderson, J. S. & Parker, R. P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J* **17**, 1497-506.

- Atzorn, V., Fragapane, P. & Kiss, T. (2004). U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. *Mol Cell Biol* **24**, 1769-78.
- Azum-Gelade, M. C., Noaillac-Depeyre, J., Caizergues-Ferrer, M. & Gas, N. (1994). Cell cycle redistribution of U3 snRNA and fibrillarin. Presence in the cytoplasmic nucleolus remnant and in the prenucleolar bodies at telophase. *J Cell Sci* **107 (Pt 2)**, 463-75.
- Baccon, J., Pellizzoni, L., Rappsilber, J., Mann, M. & Dreyfuss, G. (2002). Identification and characterization of Gemin7, a novel component of the survival of motor neuron complex. *J Biol Chem* **277**, 31957-62.
- Bachand, F., Boisvert, F. M., Cote, J., Richard, S. & Autexier, C. (2002). The product of the survival of motor neuron (SMN) gene is a human telomerase-associated protein. *Mol Biol Cell* **13**, 3192-202.
- Baker, D. L., Youssef, O. A., Chastkofsky, M. I., Dy, D. A., Terns, R. M. & Terns, M. P. (2005). RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev* **19**, 1238-48.
- Bakin, A. V. & Ofengand, J. (1998). Mapping of pseudouridine residues in RNA to nucleotide resolution. *Methods Mol Biol* **77**, 297-309.
- Balakin, A. G., Smith, L. & Fournier, M. J. (1996). The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823-34.
- Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G. & Wandless, T. J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004.

- Bardoni, B., Schenck, A. & Mandel, J. L. (1999). A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. *Hum Mol Genet* **8**, 2557-66.
- Baserga, S. J., Gilmore-Hebert, M. & Yang, X. W. (1992). Distinct molecular signals for nuclear import of the nucleolar snRNA, U3. *Genes Dev* **6**, 1120-30.
- Bassell, G. J. & Singer, R. H. (2001). Neuronal RNA localization and the cytoskeleton. *Results Probl Cell Differ* **34**, 41-56.
- Battle, D. J., Kasim, M., Wang, J. & Dreyfuss, G. (2007). SMN-independent subunits of the SMN complex. Identification of a small nuclear ribonucleoprotein assembly intermediate. *J Biol Chem* **282**, 27953-9.
- Battle, D. J., Lau, C. K., Wan, L., Deng, H., Lotti, F. & Dreyfuss, G. (2006). The Gemin5 protein of the SMN complex identifies snRNAs. *Mol Cell* **23**, 273-9.
- Beltrame, M. & Tollervey, D. (1992). Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J* **11**, 1531-42.
- Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y. & Bentwich, Z. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* **37**, 766-70.
- Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R. H. & Cuppen, E. (2005). Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120**, 21-4.
- Bernstein, K. A., Gallagher, J. E., Mitchell, B. M., Granneman, S. & Baserga, S. J. (2004). The small-subunit processome is a ribosome assembly intermediate. *Eukaryot Cell* **3**, 1619-26.

- Bertrand, S., Burlet, P., Clermont, O., Huber, C., Fondrat, C., Thierry-Mieg, D., Munnich, A. & Lefebvre, S. (1999). The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. *Hum Mol Genet* **8**, 775-82.
- Billy, E., Wegierski, T., Nasr, F. & Filipowicz, W. (2000). Rcl1p, the yeast protein similar to the RNA 3'-phosphate cyclase, associates with U3 snoRNP and is required for 18S rRNA biogenesis. *EMBO J* **19**, 2115-26.
- Boulon, S., Marmier-Gourrier, N., Pradet-Balade, B., Wurth, L., Verheggen, C., Jady, B. E., Rothe, B., Pescia, C., Robert, M. C., Kiss, T., Bardoni, B., Krol, A., Branlant, C., Allmang, C., Bertrand, E. & Charpentier, B. (2008). The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J Cell Biol* **180**, 579-95.
- Boulon, S., Verheggen, C., Jady, B. E., Girard, C., Pescia, C., Paul, C., Ospina, J. K., Kiss, T., Matera, A. G., Bordonne, R. & Bertrand, E. (2004). PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol Cell* **16**, 777-87.
- Brahms, H., Meheus, L., de Brabandere, V., Fischer, U. & Luhrmann, R. (2001). Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA* **7**, 1531-42.
- Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R., Davies, K. E., Leppert, M., Ziter, F., Wood, D. & et al. (1990). Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3. *Nature* **344**, 540-1.
- Buhler, D., Raker, V., Luhrmann, R. & Fischer, U. (1999). Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. *Hum Mol Genet* **8**, 2351-7.

- Burkard, K. T. & Butler, J. S. (2000). A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p. *Mol Cell Biol* **20**, 604-16.
- Bussaglia, E., Clermont, O., Tizzano, E., Lefebvre, S., Burglen, L., Cruaud, C., Urtizberea, J. A., Colomer, J., Munnich, A., Baiget, M. & et al. (1995). A frame-shift deletion in the survival motor neuron gene in Spanish spinal muscular atrophy patients. *Nat Genet* **11**, 335-7.
- Cabart, P., Chew, H. K. & Murphy, S. (2004). BRCA1 cooperates with NUFIP and P-TEFb to activate transcription by RNA polymerase II. *Oncogene* **23**, 5316-29.
- Caffarelli, E., Fatica, A., Prislei, S., De Gregorio, E., Fragapane, P. & Bozzoni, I. (1996). Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *EMBO J* **15**, 1121-31.
- Cahill, N. M., Friend, K., Speckmann, W., Li, Z. H., Terns, R. M., Terns, M. P. & Steitz, J. A. (2002). Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J* **21**, 3816-28.
- Carissimi, C., Baccon, J., Straccia, M., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J. & Pellizzoni, L. (2005). Unrip is a component of SMN complexes active in snRNP assembly. *FEBS Lett* **579**, 2348-54.
- Carissimi, C., Saieva, L., Gabanella, F. & Pellizzoni, L. (2006). Gemin8 is required for the architecture and function of the survival motor neuron complex. *J Biol Chem* **281**, 37009-16.
- Carmo-Fonseca, M., Pepperkok, R., Carvalho, M. T. & Lamond, A. I. (1992). Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *J Cell Biol* **117**, 1-14.

- Carvalho, T., Almeida, F., Calapez, A., Lafarga, M., Berciano, M. T. & Carmo-Fonseca, M. (1999). The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. *J Cell Biol* **147**, 715-28.
- Chan, Y. B., Miguel-Aliaga, I., Franks, C., Thomas, N., Trulzsch, B., Sattelle, D. B., Davies, K. E. & van den Heuvel, M. (2003). Neuromuscular defects in a *Drosophila* survival motor neuron gene mutant. *Hum Mol Genet* **12**, 1367-76.
- Chanfreau, G., Legrain, P. & Jacquier, A. (1998). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J Mol Biol* **284**, 975-88.
- Chari, A., Golas, M. M., Klingenhager, M., Neuenkirchen, N., Sander, B., Englbrecht, C., Sickmann, A., Stark, H. & Fischer, U. (2008). An assembly chaperone collaborates with the SMN complex to generate spliceosomal snRNPs. *Cell* **135**, 497-509.
- Charpentier, B., Muller, S. & Branlant, C. (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* **33**, 3133-44.
- Charroux, B., Pellizzoni, L., Perkinson, R. A., Shevchenko, A., Mann, M. & Dreyfuss, G. (1999). Gemin3: A novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems. *J Cell Biol* **147**, 1181-94.
- Charroux, B., Pellizzoni, L., Perkinson, R. A., Yong, J., Shevchenko, A., Mann, M. & Dreyfuss, G. (2000). Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli. *J Cell Biol* **148**, 1177-86.

- Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tamma, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D. K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D. S. & Gingeras, T. R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149-54.
- Chooi, W. Y. & Leiby, K. R. (1981). An electron microscopic method for localization of ribosomal proteins during transcription of ribosomal DNA: a method for studying protein assembly. *Proc Natl Acad Sci U S A* **78**, 4823-7.
- Chu, S., Archer, R. H., Zengel, J. M. & Lindahl, L. (1994). The RNA of RNase MRP is required for normal processing of ribosomal RNA. *Proc Natl Acad Sci U S A* **91**, 659-63.
- Cifuentes-Diaz, C., Frugier, T., Tiziano, F. D., Lacene, E., Roblot, N., Joshi, V., Moreau, M. H. & Melki, J. (2001). Deletion of murine SMN exon 7 directed to skeletal muscle leads to severe muscular dystrophy. *J Cell Biol* **152**, 1107-14.
- Cifuentes-Diaz, C., Nicole, S., Velasco, M. E., Borra-Cebrian, C., Panozzo, C., Frugier, T., Millet, G., Roblot, N., Joshi, V. & Melki, J. (2002). Neurofilament accumulation at the motor endplate and lack of axonal sprouting in a spinal muscular atrophy mouse model. *Hum Mol Genet* **11**, 1439-47.
- Colley, A., Beggs, J. D., Tollervy, D. & Lafontaine, D. L. (2000). Dhr1p, a putative DEAH-box RNA helicase, is associated with the box C+D snoRNP U3. *Mol Cell Biol* **20**, 7238-46.
- Collins, C. A. & Guthrie, C. (2000). The question remains: is the spliceosome a ribozyme? *Nat Struct Biol* **7**, 850-4.
- Currie, D. A. & Bate, M. (1995). Innervation is essential for the development and differentiation of a sex-specific adult muscle in *Drosophila melanogaster*. *Development* **121**, 2549-57.

- Dai, M. S. & Lu, H. (2008). Crosstalk between c-Myc and ribosome in ribosomal biogenesis and cancer. *J Cell Biochem* **105**, 670-7.
- Darzacq, X., Jady, B. E., Verheggen, C., Kiss, A. M., Bertrand, E. & Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J* **21**, 2746-56.
- Darzacq, X. & Kiss, T. (2000). Processing of intron-encoded box C/D small nucleolar RNAs lacking a 5',3'-terminal stem structure. *Mol Cell Biol* **20**, 4522-31.
- Darzacq, X., Kittur, N., Roy, S., Shav-Tal, Y., Singer, R. H. & Meier, U. T. (2008). Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J Cell Biol* **173**, 207-18.
- Das, R. & Reed, R. (1999). Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA* **5**, 1504-8.
- Decatur, W. A. & Fournier, M. J. (2002). rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-51.
- Dez, C., Froment, C., Noaillac-Depeyre, J., Monsarrat, B., Caizergues-Ferrer, M. & Henry, Y. (2004). Npa1p, a component of very early pre-60S ribosomal particles, associates with a subset of small nucleolar RNPs required for peptidyl transferase center modification. *Mol Cell Biol* **24**, 6324-37.
- Dong, X. Y., Rodriguez, C., Guo, P., Sun, X., Talbot, J. T., Zhou, W., Petros, J., Li, Q., Vessella, R. L., Kibel, A. S., Stevens, V. L., Calle, E. E. & Dong, J. T. (2008). SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum Mol Genet* **17**, 1031-42.

- Donmez, G., Hartmuth, K. & Luhrmann, R. (2004). Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA* **10**, 1925-33.
- Dragon, F., Gallagher, J. E., Compagnone-Post, P. A., Mitchell, B. M., Porwancher, K. A., Wehner, K. A., Wormsley, S., Settlage, R. E., Shabanowitz, J., Osheim, Y., Beyer, A. L., Hunt, D. F. & Baserga, S. J. (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967-70.
- Dunbar, D. A., Wormsley, S., Agentis, T. M. & Baserga, S. J. (1997). Mpp10p, a U3 small nucleolar ribonucleoprotein component required for pre-18S rRNA processing in yeast. *Mol Cell Biol* **17**, 5803-12.
- Dworniczak, B. & Mirault, M. E. (1987). Structure and expression of a human gene coding for a 71 kd heat shock 'cognate' protein. *Nucleic Acids Res* **15**, 5181-97.
- Eggert, C., Chari, A., Lagerbauer, B. & Fischer, U. (2006). Spinal muscular atrophy: the RNP connection. *Trends Mol Med* **12**, 113-21.
- Elbashir, S. M., Harborth, J., Weber, K. & Tuschl, T. (2002). Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* **26**, 199-213.
- Esguerra, J., Warringer, J. & Blomberg, A. (2008). Functional importance of individual rRNA 2'-O-ribose methylations revealed by high-resolution phenotyping. *RNA* **14**, 649-56.
- Fatica, A., Oeffinger, M., Dlakic, M. & Tollervey, D. (2003). Nob1p is required for cleavage of the 3' end of 18S rRNA. *Mol Cell Biol* **23**, 1798-807.

- Feng, W., Gubitz, A. K., Wan, L., Battle, D. J., Dostie, J., Golembe, T. J. & Dreyfuss, G. (2005). Gemins modulate the expression and activity of the SMN complex. *Hum Mol Genet* **14**, 1605-11.
- Filipowicz, W., Pelczar, P., Pogacic, V. & Dragon, F. (1999). Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification. *Acta Biochim Pol* **46**, 377-89.
- Flygare, J. & Karlsson, S. (2007). Diamond-Blackfan anemia: erythropoiesis lost in translation. *Blood* **109**, 3152-4.
- Fournier, M. J. & Maxwell, E. S. (1993). The nucleolar snRNAs: catching up with the spliceosomal snRNAs. *Trends Biochem Sci* **18**, 131-5.
- Frey, M. R., Bailey, A. D., Weiner, A. M. & Matera, A. G. (1999). Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. *Curr Biol* **9**, 126-35.
- Frey, M. R. & Matera, A. G. (2001). RNA-mediated interaction of Cajal bodies and U2 snRNA genes. *J Cell Biol* **154**, 499-509.
- Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A. & Dreyfuss, G. (2001). SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Mol Cell* **7**, 1111-7.
- Frilander, M. J. & Steitz, J. A. (2001). Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Mol Cell* **7**, 217-26.
- Frilander, M. J. & Steitz, J. A. (2001). Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Mol Cell* **7**, 217-26.

- Fu, X. D. (1995). The superfamily of arginine/serine-rich splicing factors. *RNA* **1**, 663-80.
- Fury, M. G. & Zieve, G. W. (1996). U6 snRNA maturation and stability. *Exp Cell Res* **228**, 160-3.
- Gadad, O., Strauss, D., Braspenning, J., Hoepfner, D., Petfalski, E., Philippsen, P., Tollervey, D. & Hurt, E. (2001). A nuclear AAA-type ATPase (Rix7p) is required for biogenesis and nuclear export of 60S ribosomal subunits. *EMBO J* **20**, 3695-704.
- Galardi, S., Fatica, A., Bachi, A., Scaloni, A., Presutti, C. & Bozzoni, I. (2002). Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation of target RNA *in vitro*. *Mol Cell Biol* **22**, 6663-8.
- Gallagher, J. E., Dunbar, D. A., Granneman, S., Mitchell, B. M., Osheim, Y., Beyer, A. L. & Baserga, S. J. (2004). RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. *Genes Dev* **18**, 2506-17.
- Ganot, P., Bortolin, M. L. & Kiss, T. (1997). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799-809.
- Ganot, P., Caizergues-Ferrer, M. & Kiss, T. (1997). The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11**, 941-56.
- Ganot, P., Jady, B. E., Bortolin, M. L., Darzacq, X. & Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol* **19**, 6906-17.

- Gao, L., Frey, M. R. & Matera, A. G. (1997). Human genes encoding U3 snRNA associate with coiled bodies in interphase cells and are clustered on chromosome 17p11.2 in a complex inverted repeat structure. *Nucleic Acids Res* **25**, 4740-7.
- Gautier, T., Berges, T., Tollervey, D. & Hurt, E. (1997). Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol* **17**, 7088-98.
- Gavrillov, D. K., Shi, X., Das, K., Gilliam, T. C. & Wang, C. H. (1998). Differential SMN2 expression associated with SMA severity. *Nat Genet* **20**, 230-1.
- Gieseemann, T., Rathke-Hartlieb, S., Rothkegel, M., Bartsch, J. W., Buchmeier, S., Jockusch, B. M. & Jockusch, H. (1999). A role for polyproline motifs in the spinal muscular atrophy protein SMN. Profilins bind to and colocalize with smn in nuclear gems. *J Biol Chem* **274**, 37908-14.
- Giorgi, C., Fatica, A., Nagel, R. & Bozzoni, I. (2001). Release of U18 snoRNA from its host intron requires interaction of Nop1p with the Rnt1p endonuclease. *EMBO J* **20**, 6856-65.
- Girard, C., Verheggen, C., Neel, H., Cammas, A., Vagner, S., Soret, J., Bertrand, E. & Bordonne, R. (2008). Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *J Biol Chem* **283**, 2060-9.
- Gleizes, P. E., Noaillac-Depeyre, J., Leger-Silvestre, I., Teulieres, F., Dauxois, J. Y., Pommet, D., Azum-Gelade, M. C. & Gas, N. (2001). Ultrastructural localization of rRNA shows defective nuclear export of preribosomes in mutants of the Nup82p complex. *J Cell Biol* **155**, 923-36.

- Glibetic, M., Larson, D. E., Sienna, N., Bachellerie, J. P. & Sells, B. H. (1992). Regulation of U3 snRNA expression during myoblast differentiation. *Exp Cell Res* **202**, 183-9.
- Gonzales, F. A., Zanchin, N. I., Luz, J. S. & Oliveira, C. C. (2005). Characterization of *Saccharomyces cerevisiae* Nop17p, a novel Nop58p-interacting protein that is involved in Pre-rRNA processing. *J Mol Biol* **346**, 437-55.
- Granneman, S. & Baserga, S. J. (2004). Ribosome biogenesis: of knobs and RNA processing. *Exp Cell Res* **296**, 43-50.
- Granneman, S., Gallagher, J. E., Vogelzangs, J., Horstman, W., van Venrooij, W. J., Baserga, S. J. & Pruijn, G. J. (2003). The human Imp3 and Imp4 proteins form a ternary complex with hMpp10, which only interacts with the U3 snoRNA in 60-80S ribonucleoprotein complexes. *Nucleic Acids Res* **31**, 1877-87.
- Granneman, S., Pruijn, G. J., Horstman, W., van Venrooij, W. J., Luhrmann, R. & Watkins, N. J. (2002). The hU3-55K protein requires 15.5K binding to the box B/C motif as well as flanking RNA elements for its association with the U3 small nucleolar RNA *in Vitro*. *J Biol Chem* **277**, 48490-500.
- Granneman, S., Vogelzangs, J., Luhrmann, R., van Venrooij, W. J., Pruijn, G. J. & Watkins, N. J. (2004). Role of pre-rRNA base pairing and 80S complex formation in subnucleolar localization of the U3 snoRNP. *Mol Cell Biol* **24**, 8600-10.
- Grummt, I. & Pikaard, C. S. (2003). Epigenetic silencing of RNA polymerase I transcription. *Nat Rev Mol Cell Biol* **4**, 641-9.
- Gubitz, A. K., Mourelatos, Z., Abel, L., Rappsilber, J., Mann, M. & Dreyfuss, G. (2002). Gemin5, a novel WD repeat protein component of the SMN complex that binds Sm proteins. *J Biol Chem* **277**, 5631-6.

- Hadjiolova, K. V., Nicoloso, M., Mazan, S., Hadjiolov, A. A. & Bachellerie, J. P. (1993). Alternative pre-rRNA processing pathways in human cells and their alteration by cycloheximide inhibition of protein synthesis. *Eur J Biochem* **212**, 211-5.
- Hamma, T., Reichow, S. L., Varani, G. & Ferre-D'Amare, A. R. (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol* **12**, 1101-7.
- Hausmann, S. & Shuman, S. (2005). Specificity and mechanism of RNA cap guanine-N2 methyltransferase (Tgs1). *J Biol Chem* **280**, 4021-4.
- Hayano, T., Yanagida, M., Yamauchi, Y., Shinkawa, T., Isobe, T. & Takahashi, N. (2003). Proteomic analysis of human Nop56p-associated pre-ribosomal ribonucleoprotein complexes. Possible link between Nop56p and the nucleolar protein treacle responsible for Treacher Collins syndrome. *J Biol Chem* **278**, 34309-19.
- He, W. & Parker, R. (2000). Functions of Lsm proteins in mRNA degradation and splicing. *Curr Opin Cell Biol* **12**, 346-50.
- Hebert, M. D., Shpargel, K. B., Ospina, J. K., Tucker, K. E. & Matera, A. G. (2002). Coilin methylation regulates nuclear body formation. *Dev Cell* **3**, 329-37.
- Heiss, N. S., Knight, S. W., Vulliamey, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., Poustka, A. & Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* **19**, 32-8.

- Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J. P. & Caizergues-Ferrer, M. (1998). Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J* **17**, 7078-90.
- Henras, A. K., Capeyrou, R., Henry, Y. & Caizergues-Ferrer, M. (2004). Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs. *RNA* **10**, 1704-12.
- Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H. & Luhrmann, R. (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J* **14**, 2076-88.
- Hernandez, N. (2001). Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem* **276**, 26733-6.
- Hiley, S. L., Babak, T. & Hughes, T. R. (2005). Global analysis of yeast RNA processing identifies new targets of RNase III and uncovers a link between tRNA 5' end processing and tRNA splicing. *Nucleic Acids Res* **33**, 3048-56.
- Hirose, T. & Steitz, J. A. (2001). Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc Natl Acad Sci U S A* **98**, 12914-9.
- Huang, S. (2000). Review: Perinucleolar Structures. *J Structural Biology* **129**, 233-240.
- Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M. & Luhrmann, R. (1998). Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J* **17**, 4114-26.

- Huber, J., Dickmanns, A. & Luhrmann, R. (2002). The importin-beta binding domain of snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs *in vitro*. *J Cell Biol* **156**, 467-79.
- Hughes, J. M. (1996). Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. *J Mol Biol* **259**, 645-54.
- Hughes, J. M. & Ares, M., Jr. (1991). Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J* **10**, 4231-9.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* **98**, 4569-74.
- Jablonka, S., Bandilla, M., Wiese, S., Buhler, D., Wirth, B., Sendtner, M. & Fischer, U. (2001). Co-regulation of survival of motor neuron (SMN) protein and its interactor SIP1 during development and in spinal muscular atrophy. *Hum Mol Genet* **10**, 497-505.
- Jady, B. E., Darzacq, X., Tucker, K. E., Matera, A. G., Bertrand, E. & Kiss, T. (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J* **22**, 1878-88.
- Jady, B. E. & Kiss, T. (2001). A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J* **20**, 541-51.
- Jarmolowski, A., Zagorski, J., Li, H. V. & Fournier, M. J. (1990). Identification of essential elements in U14 RNA of *Saccharomyces cerevisiae*. *EMBO J* **9**, 4503-9.

- Jones, K. W., Gorzynski, K., Hales, C. M., Fischer, U., Badbanchi, F., Terns, R. M. & Terns, M. P. (2001). Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. *J Biol Chem* **276**, 38645-51.
- Kass, S., Craig, N. & Sollner-Webb, B. (1987). Primary processing of mammalian rRNA involves two adjacent cleavages and is not species specific. *Mol Cell Biol* **7**, 2891-8.
- Kass, S., Tyc, K., Steitz, J. A. & Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell* **60**, 897-908.
- Kenmochi, N., Higa, S., Yoshihama, M. & Tanaka, T. (1996). U14 snoRNAs are encoded in introns of human ribosomal protein S13 gene. *Biochem Biophys Res Commun* **228**, 371-4.
- King, T. H., Decatur, W. A., Bertrand, E., Maxwell, E. S. & Fournier, M. J. (2001). A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol Cell Biol* **21**, 7731-46.
- King, T. H., Liu, B., McCully, R. R. & Fournier, M. J. (2003). Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* **11**, 425-35.
- Kiss, A. M., Jady, B. E., Darzacq, X., Verheggen, C., Bertrand, E. & Kiss, T. (2002). A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Res* **30**, 4643-9.
- Kiss, T. (2004). Biogenesis of small nuclear RNPs. *J Cell Sci* **117**, 5949-51.

- Kiss, T. & Filipowicz, W. (1995). Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev* **9**, 1411-24.
- Kiss-Laszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M. & Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**, 1077-88.
- Kiss-Laszlo, Z., Henry, Y. & Kiss, T. (1998). Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J* **17**, 797-807.
- Kitao, S., Segref, A., Kast, J., Wilm, M., Mattaj, I. W. & Ohno, M. (2008). A compartmentalized phosphorylation/dephosphorylation system that regulates U snRNA export from the nucleus. *Mol Cell Biol* **28**, 487-97.
- Klemm, R. D., Goodrich, J. A., Zhou, S. & Tjian, R. (1995). Molecular cloning and expression of the 32-kDa subunit of human TFIID reveals interactions with VP16 and TFIIB that mediate transcriptional activation. *Proc Natl Acad Sci U S A* **92**, 5788-92.
- Konarska, M. M. & Sharp, P. A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* **49**, 763-74.
- Koonin, E. V., Bork, P. & Sander, C. (1994). A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Res* **22**, 2166-7.
- Koonin, E. V., Bork, P. & Sander, C. (1994). A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Res* **22**, 2166-7.
- Krogan, N. J., Peng, W. T., Cagney, G., Robinson, M. D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D. P., Beattie, B. K., Lalev, A.,

- Zhang, W., Davierwala, A. P., Mnaimneh, S., Starostine, A., Tikuisis, A. P., Grigull, J., Datta, N., Bray, J. E., Hughes, T. R., Emili, A. & Greenblatt, J. F. (2004). High-definition macromolecular composition of yeast RNA-processing complexes. *Mol Cell* **13**, 225-39.
- Kroiss, M., Schultz, J., Wiesner, J., Chari, A., Sickmann, A. & Fischer, U. (2008). Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **105**, 10045-50.
 - Kufel, J., Allmang, C., Chanfreau, G., Petfalski, E., Lafontaine, D. L. & Tollervey, D. (2000). Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Mol Cell Biol* **20**, 5415-24.
 - Kufel, J., Allmang, C., Verdone, L., Beggs, J. & Tollervey, D. (2003). A complex pathway for 3' processing of the yeast U3 snoRNA. *Nucleic Acids Res* **31**, 6788-97.
 - Kufel, J., Dichtl, B. & Tollervey, D. (1999). Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA* **5**, 909-17.
 - Kuhn, J. F., Tran, E. J. & Maxwell, E. S. (2002). Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res* **30**, 931-41.
 - Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. & Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev* **12**, 527-37.
 - Lafontaine, D. L. & Tollervey, D. (1999). Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA* **5**, 455-67.

- Lafontaine, D. L. & Tollervey, D. (2000). Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol Cell Biol* **20**, 2650-9.
- Lafontaine, D. L. & Tollervey, D. (2001). The function and synthesis of ribosomes. *Nat Rev Mol Cell Biol* **2**, 514-20.
- Lakkaraju, A. K., Mary, C., Scherrer, A., Johnson, A. E. & Strub, K. (2008). SRP keeps polypeptides translocation-competent by slowing translation to match limiting ER-targeting sites. *Cell* **133**, 440-51.
- Lamond, A. I. & Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4**, 605-12.
- Laneve, P., Altieri, F., Fiori, M. E., Scaloni, A., Bozzoni, I. & Caffarelli, E. (2003). Purification, cloning, and characterization of XendoU, a novel endoribonuclease involved in processing of intron-encoded small nucleolar RNAs in *Xenopus laevis*. *J Biol Chem* **278**, 13026-32.
- Lange, T. S. & Gerbi, S. A. (2000). Transient nucleolar localization Of U6 small nuclear RNA in *Xenopus Laevis* oocytes. *Mol Biol Cell* **11**, 2419-28.
- Lee, C. Y., Lee, A. & Chanfreau, G. (2003). The roles of endonucleolytic cleavage and exonucleolytic digestion in the 5'-end processing of *S. cerevisiae* box C/D snoRNAs. *RNA* **9**, 1362-70.
- Lee, S. J. & Baserga, S. J. (1999). Imp3p and Imp4p, two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. *Mol Cell Biol* **19**, 5441-52.
- Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H. & Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-60.

- Lemm, I., Girard, C., Kuhn, A. N., Watkins, N. J., Schneider, M., Bordonne, R. & Luhrmann, R. (2006). Ongoing U snRNP biogenesis is required for the integrity of Cajal bodies. *Mol Biol Cell* **17**, 3221-31.
- Lestrade, L. & Weber, M. J. (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* **34**, D158-62.
- Leung, A. K., Gerlich, D., Miller, G., Lyon, C., Lam, Y. W., Lleres, D., Daigle, N., Zomerdijs, J., Ellenberg, J. & Lamond, A. I. (2004). Quantitative kinetic analysis of nucleolar breakdown and reassembly during mitosis in live human cells. *J Cell Biol* **166**, 787-800.
- Li, H. D., Zagorski, J. & Fournier, M. J. (1990). Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* **10**, 1145-52.
- Li, L. & Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**, 302-7.
- Liang, W. Q. & Fournier, M. J. (1995). U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Genes Dev* **9**, 2433-43.
- Libri, D., Dower, K., Boulay, J., Thomsen, R., Rosbash, M. & Jensen, T. H. (2002). Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Mol Cell Biol* **22**, 8254-66.
- Lin, X., Yang, J., Chen, J., Gunasekera, A., Fesik, S. W. & Shen, Y. (2004). Development of a tightly regulated U6 promoter for shRNA expression. *FEBS Lett* **577**, 376-80.

- Liu, J. & Maxwell, E. S. (1990). Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. *Nucleic Acids Res* **18**, 6565-71.
- Liu, J., Valencia-Sanchez, M. A., Hannon, G. J. & Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* **7**, 719-23.
- Liu, Q. & Dreyfuss, G. (1996). A novel nuclear structure containing the survival of motor neurons protein. *EMBO J* **15**, 3555-65.
- Liu, Q., Fischer, U., Wang, F. & Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell* **90**, 1013-21.
- Liu, Q., Greimann, J. C. & Lima, C. D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223-37.
- Lorson, C. L., Strasswimmer, J., Yao, J. M., Baleja, J. D., Hahnen, E., Wirth, B., Le, T., Burghes, A. H. & Androphy, E. J. (1998). SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nat Genet* **19**, 63-6.
- Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R. & Golub, T. R. (2005). MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-8.
- Luo, H. R., Moreau, G. A., Levin, N. & Moore, M. J. (1999). The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA* **5**, 893-908.
- Lygerou, Z., Allmang, C., Tollervey, D. & Seraphin, B. (1996). Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science* **272**, 268-70.

- Lyman, S. K., Gerace, L. & Baserga, S. J. (1999). Human Nop5/Nop58 is a component common to the box C/D small nucleolar ribonucleoproteins. *RNA* **5**, 1597-604.
- Maden, B. E. (1990). The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol* **39**, 241-303.
- Makarova, O. V., Makarov, E. M., Liu, S., Vornlocher, H. P. & Luhrmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J* **21**, 1148-57.
- Manival, X., Charron, C., Fourmann, J. B., Godard, F., Charpentier, B. & Branlant, C. (2006). Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res* **34**, 826-39.
- Maroney, P. A., Romfo, C. M. & Nilsen, T. W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol Cell* **6**, 317-28.
- Marshallsay, C. & Luhrmann, R. (1994). *In vitro* nuclear import of snRNPs: cytosolic factors mediate m3G-cap dependence of U1 and U2 snRNP transport. *EMBO J* **13**, 222-31.
- Matera, A. G. & Frey, M. R. (1998). Coiled bodies and gems: Janus or gemini? *Am J Hum Genet* **63**, 317-21.
- Maul, G. G., Negorev, D., Bell, P. & Ishov, A. M. (2000). Review: properties and assembly mechanisms of ND10, PML bodies, or PODs. *J Struct Biol* **129**, 278-87.

- Mazan, S., Mattei, M. G., Roeckel, N., Qu, L. H. & Bachellerie, J. P. (1993). In humans all U3 genes map to chromosome 17p12-->p11, but in mouse the U3A and U3B genes are located on different chromosomes. *Cytogenet Cell Genet* **62**, 203-6.
- McKeegan, K. S., Debieux, C. M., Boulon, S., Bertrand, E. & Watkins, N. J. (2007). A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol Cell Biol* **27**, 6782-93.
- McWhorter, M. L., Monani, U. R., Burghes, A. H. & Beattie, C. E. (2003). Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. *J Cell Biol* **162**, 919-31.
- Medlin, J. E., Uguen, P., Taylor, A., Bentley, D. L. & Murphy, S. (2003). The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO J* **22**, 925-34.
- Meier, U. T. & Blobel, G. (1992). Nopp140 shuttles on tracks between nucleolus and cytoplasm. *Cell* **70**, 127-38.
- Meier, U. T. & Blobel, G. (1994). NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *J Cell Biol* **127**, 1505-14.
- Meister, G., Buhler, D., Pillai, R., Lottspeich, F. & Fischer, U. (2001). A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat Cell Biol* **3**, 945-9.
- Meister, G., Eggert, C. & Fischer, U. (2002). SMN-mediated assembly of RNPs: a complex story. *Trends Cell Biol* **12**, 472-8.

- Menichelli, E., Isel, C., Oubridge, C. & Nagai, K. (2007). Protein-induced conformational changes of RNA during the assembly of human signal recognition particle. *J Mol Biol* **367**, 187-203.
- Mereau, A., Fournier, R., Gregoire, A., Mougin, A., Fabrizio, P., Luhrmann, R. & Branlant, C. (1997). An *in vivo* and *in vitro* structure-function analysis of the *Saccharomyces cerevisiae* U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. *J Mol Biol* **273**, 552-71.
- Michaud, N. & Goldfarb, D. (1992). Microinjected U snRNAs are imported to oocyte nuclei via the nuclear pore complex by three distinguishable targeting pathways. *J Cell Biol* **116**, 851-61.
- Michot, B., Joseph, N., Mazan, S. & Bachellerie, J. P. (1999). Evolutionarily conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with the snoRNA U8 detected by comparative analysis of new mouse sequences. *Nucleic Acids Res* **27**, 2271-82.
- Misteli, T., Caceres, J. F. & Spector, D. L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523-7.
- Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervy, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell* **91**, 457-66.
- Mitchell, P., Petfalski, E. & Tollervy, D. (1996). The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev* **10**, 502-13.
- Mitrousis, G., Olia, A. S., Walker-Kopp, N. & Cingolani, G. (2008). Molecular basis for the recognition of snurportin 1 by importin beta. *J Biol Chem* **283**, 7877-84.

- Monani, U. R., Sendtner, M., Coover, D. D., Parsons, D. W., Andreassi, C., Le, T. T., Jablonka, S., Schrank, B., Rossoll, W., Prior, T. W., Morris, G. E. & Burghes, A. H. (2000). The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in *Smn*(^{-/-}) mice and results in a mouse with spinal muscular atrophy. *Hum Mol Genet* **9**, 333-9.
- Morrissey, J. P. & Tollervey, D. (1993). Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Mol Cell Biol* **13**, 2469-77.
- Morrissey, J. P. & Tollervey, D. (1997). U14 small nucleolar RNA makes multiple contacts with the pre-ribosomal RNA. *Chromosoma* **105**, 515-22.
- Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A. G., Bertrand, E., Tazi, J. & Bordonne, R. (2003). Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep* **4**, 616-22.
- Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. & Bordonne, R. (2002). Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol Cell* **9**, 891-901.
- Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. & Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**, 720-8.
- Muqit, M. M., Moss, J., Sewry, C. & Lane, R. J. (2004). Phenotypic variability in siblings with type III spinal muscular atrophy. *J Neurol Neurosurg Psychiatry* **75**, 1762-4.
- Murdoch, K. J. & Allison, L. A. (1996). A role for ribosomal protein L5 in the nuclear import of 5S rRNA in *Xenopus* oocytes. *Exp Cell Res* **227**, 332-43.

- Narayanan, A., Speckmann, W., Terns, R. & Terns, M. P. (1999). Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol Biol Cell* **10**, 2131-47.
- Narayanan, U., Achsel, T., Luhrmann, R. & Matera, A. G. (2004). Coupled *in vitro* import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol Cell* **16**, 223-34.
- Narayanan, U., Ospina, J. K., Frey, M. R., Hebert, M. D. & Matera, A. G. (2002). SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. *Hum Mol Genet* **11**, 1785-95.
- Nesic, D., Tanackovic, G. & Kramer, A. (2004). A role for Cajal bodies in the final steps of U2 snRNP biogenesis. *J Cell Sci* **117**, 4423-33.
- Newman, A. J., Teigelkamp, S. & Beggs, J. D. (1995). snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA* **1**, 968-80.
- Newman, D. R., Kuhn, J. F., Shanab, G. M. & Maxwell, E. S. (2000). Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* **6**, 861-79.
- Ni, J., Tien, A. L. & Fournier, M. J. (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565-73.
- Nissan, T. A., Bassler, J., Petfalski, E., Tollervey, D. & Hurt, E. (2002). 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. *EMBO J* **21**, 5539-47.
- Normand, C., Capeyrou, R., Quevillon-Cheruel, S., Mougin, A., Henry, Y. & Caizergues-Ferrer, M. (2006). Analysis of the binding of the N-terminal conserved domain of yeast Cbf5p to a box H/ACA snoRNA. *RNA* **12**, 1868-82.

- Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R. & Luhrmann, R. (1999). Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J* **18**, 6119-33.
- Nottrott, S., Urlaub, H. & Luhrmann, R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J* **21**, 5527-38.
- Novoyatleva, T., Heinrich, B., Tang, Y., Benderska, N., Butchbach, M. E., Lorson, C. L., Lorson, M. A., Ben-Dov, C., Fehlbaum, P., Bracco, L., Burghes, A. H., Bollen, M. & Stamm, S. (2008). Protein phosphatase 1 binds to the RNA recognition motif of several splicing factors and regulates alternative pre-mRNA processing. *Hum Mol Genet* **17**, 52-70.
- Ofengand, J. (2002). Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett* **514**, 17-25.
- Ofengand, J. & Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J Mol Biol* **266**, 246-68.
- Ohno, M., Segref, A., Bachi, A., Wilm, M. & Mattaj, I. W. (2000). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187-98.
- Omer, A. D., Lowe, T. M., Russell, A. G., Ebhardt, H., Eddy, S. R. & Dennis, P. P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science* **288**, 517-22.
- Omer, A. D., Ziesche, S., Ebhardt, H. & Dennis, P. P. (2002). *In vitro* reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci U S A* **99**, 5289-94.

- Ooi, S. L., Samarsky, D. A., Fournier, M. J. & Boeke, J. D. (1998). Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA* **4**, 1096-110.
- Ospina, J. K., Gonsalvez, G. B., Bednenko, J., Darzynkiewicz, E., Gerace, L. & Matera, A. G. (2005). Cross-talk between snurportin1 subdomains. *Mol Biol Cell* **16**, 4660-71.
- Otter, S., Grimmmler, M., Neuenkirchen, N., Chari, A., Sickmann, A. & Fischer, U. (2007). A comprehensive interaction map of the human survival of motor neuron (SMN) complex. *J Biol Chem* **282**, 5825-33.
- Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-8.
- Oulton, R. & Harrington, L. (2000). Telomeres, telomerase, and cancer: life on the edge of genomic stability. *Curr Opin Oncol* **12**, 74-81.
- Palacios, I., Hetzer, M., Adam, S. A. & Mattaj, I. W. (1997). Nuclear import of U snRNPs requires importin beta. *EMBO J* **16**, 6783-92.
- Pannone, B. K., Xue, D. & Wolin, S. L. (1998). A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J* **17**, 7442-53.
- Paraskeva, E., Izaurralde, E., Bischoff, F. R., Huber, J., Kutay, U., Hartmann, E., Luhrmann, R. & Gorlich, D. (1999). CRM1-mediated recycling of snurportin 1 to the cytoplasm. *J Cell Biol* **145**, 255-64.
- Patel, S. B. & Bellini, M. (2008). The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **36**, 6482-93.

- Paushkin, S., Gubitz, A. K., Massenet, S. & Dreyfuss, G. (2002). The SMN complex, an assemblysome of ribonucleoproteins. *Curr Opin Cell Biol* **14**, 305-12.
- Paushkin, S., Gubitz, A. K., Massenet, S. & Dreyfuss, G. (2002). The SMN complex, an assemblysome of ribonucleoproteins. *Curr Opin Cell Biol* **14**, 305-12.
- Peculis, B. A. (1997). The sequence of the 5' end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. *Mol Cell Biol* **17**, 3702-13.
- Peculis, B. A. (2001). snoRNA nuclear import and potential for cotranscriptional function in pre-rRNA processing. *RNA* **7**, 207-19.
- Peculis, B. A. & Steitz, J. A. (1993). Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the *Xenopus* oocyte. *Cell* **73**, 1233-45.
- Peculis, B. A. & Steitz, J. A. (1994). Sequence and structural elements critical for U8 snRNP function in *Xenopus* oocytes are evolutionarily conserved. *Genes Dev* **8**, 2241-55.
- Pellizzoni, L., Baccon, J., Charroux, B. & Dreyfuss, G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Curr Biol* **11**, 1079-88.
- Pellizzoni, L., Baccon, J., Rappsilber, J., Mann, M. & Dreyfuss, G. (2002). Purification of native survival of motor neurons complexes and identification of Gemin6 as a novel component. *J Biol Chem* **277**, 7540-5.
- Pellizzoni, L., Charroux, B. & Dreyfuss, G. (1999). SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. *Proc Natl Acad Sci U S A* **96**, 11167-72.

- Pellizzoni, L., Yong, J. & Dreyfuss, G. (2002). Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298**, 1775-9.
- Peng, W. T., Robinson, M. D., Mnaimneh, S., Krogan, N. J., Cagney, G., Morris, Q., Davierwala, A. P., Grigull, J., Yang, X., Zhang, W., Mitsakakis, N., Ryan, O. W., Datta, N., Jojic, V., Pal, C., Canadien, V., Richards, D., Beattie, B., Wu, L. F., Altschuler, S. J., Roweis, S., Frey, B. J., Emili, A., Greenblatt, J. F. & Hughes, T. R. (2003). A panoramic view of yeast noncoding RNA processing. *Cell* **113**, 919-33.
- Perez-Fernandez, J., Roman, A., De Las Rivas, J., Bustelo, X. R. & Dosil, M. (2007). The 90S preribosome is a multimodular structure that is assembled through a hierarchical mechanism. *Mol Cell Biol* **27**, 5414-29.
- Petfalski, E., Dandekar, T., Henry, Y. & Tollervey, D. (1998). Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol Cell Biol* **18**, 1181-9.
- Pluk, H., Soffner, J., Luhrmann, R. & van Venrooij, W. J. (1998). cDNA cloning and characterization of the human U3 small nucleolar ribonucleoprotein complex-associated 55-kilodalton protein. *Mol Cell Biol* **18**, 488-98.
- Pombo, A., Cuello, P., Schul, W., Yoon, J. B., Roeder, R. G., Cook, P. R. & Murphy, S. (1998). Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J* **17**, 1768-78.
- Prior, T. W., Swoboda, K. J., Scott, H. D. & Hejmanowski, A. Q. (2004). Homozygous SMN1 deletions in unaffected family members and modification of the phenotype by SMN2. *Am J Med Genet A* **130A**, 307-10.

- Speckmann, W. A., Terns, R. M. & Terns, M. P. (2000). The box C/D motif directs snoRNA 5'-cap hypermethylation. *Nucleic Acids Res* **28**, 4467-73.
- Puri, T., Wendler, P., Sigala, B., Saibil, H. & Tsaneva, I. R. (2007). Dodecameric structure and ATPase activity of the human TIP48/TIP49 complex. *J Mol Biol* **366**, 179-92.
- Rajendra, T. K., Gonsalvez, G. B., Walker, M. P., Shpargel, K. B., Salz, H. K. & Matera, A. G. (2007). A *Drosophila melanogaster* model of spinal muscular atrophy reveals a function for SMN in striated muscle. *J Cell Biol* **176**, 831-41.
- Rashid, R., Aittaleb, M., Chen, Q., Spiegel, K., Demeler, B. & Li, H. (2003). Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. *J Mol Biol* **333**, 295-306.
- Rashid, R., Liang, B., Baker, D. L., Youssef, O. A., He, Y., Phipps, K., Terns, R. M., Terns, M. P. & Li, H. (2006). Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell* **21**, 249-60.
- Renvoise, B., Colasse, S., Burlet, P., Viollet, L., Meier, U. T. & Lefebvre, S. (2009). The loss of the snoRNP chaperone Nopp140 from Cajal bodies of patient fibroblasts correlates with the severity of spinal muscular atrophy. *Hum Mol Genet*.
- Renvoise, B., Khoobarry, K., Gendron, M. C., Cibert, C., Viollet, L. & Lefebvre, S. (2006). Distinct domains of the spinal muscular atrophy protein SMN are required for targeting to Cajal bodies in mammalian cells. *J Cell Sci* **119**, 680-92.
- Richard, P., Darzacq, X., Bertrand, E., Jady, B. E., Verheggen, C. & Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J* **22**, 4283-93.

- Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. & Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**, 1902-10.
- Rollenhagen, C., Muhlhauser, P., Kutay, U. & Pante, N. (2003). Importin beta-depending nuclear import pathways: role of the adapter proteins in the docking and releasing steps. *Mol Biol Cell* **14**, 2104-15.
- Rossoll, W., Jablonka, S., Andreassi, C., Kroning, A. K., Karle, K., Monani, U. R. & Sendtner, M. (2003). Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons. *J Cell Biol* **163**, 801-12.
- Rossoll, W., Kroning, A. K., Ohndorf, U. M., Steegborn, C., Jablonka, S. & Sendtner, M. (2002). Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? *Hum Mol Genet* **11**, 93-105.
- Rouquette, J., Choesmel, V. & Gleizes, P. E. (2005). Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. *EMBO J* **24**, 2862-72.
- Russell, J. & Zomerdijk, J. C. (2005). RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem Sci* **30**, 87-96.
- Saponara, A. G. & Enger, M. D. (1969). Occurrence of N²,N²,7-trimethylguanosine in minor RNA species of a mammalian cell line. *Nature* **223**, 1365-6.
- Savino, R. & Gerbi, S. A. (1990). *In vivo* disruption of *Xenopus* U3 snRNA affects ribosomal RNA processing. *EMBO J* **9**, 2299-308.

- Savino, R., Hitti, Y. & Gerbi, S. A. (1992). Genes for *Xenopus laevis* U3 small nuclear RNA. *Nucleic Acids Res* **20**, 5435-42.
- Schafer, T., Strauss, D., Petfalski, E., Tollervey, D. & Hurt, E. (2003). The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J* **22**, 1370-80.
- Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T., Luhrmann, R. (2004). RNAi knockdown of pPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J* **23**, 3000-3009.
- Scherl, A., Coute, Y., Deon, C., Calle, A., Kindbeiter, K., Sanchez, J. C., Greco, A., Hochstrasser, D. & Diaz, J. J. (2002). Functional proteomic analysis of human nucleolus. *Mol Biol Cell* **13**, 4100-9.
- Schimmang, T., Tollervey, D., Kern, H., Frank, R. & Hurt, E. C. (1989). A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J* **8**, 4015-24.
- Schmalbruch, H. & Haase, G. (2001). Spinal muscular atrophy: present state. *Brain Pathol* **11**, 231-47.
- Schmitt, M. E. & Clayton, D. A. (1993). Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* **13**, 7935-41.
- Schneider, C., Will, C. L., Makarova, O. V., Makarov, E. M. & Luhrmann, R. (2002). Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol Cell Biol* **22**, 3219-29.
- Schul, W., Groenhout, B., Koberna, K., Takagaki, Y., Jenny, A., Manders, E. M., Raska, I., van Driel, R. & de Jong, L. (1996). The RNA 3' cleavage factors CstF 64 kDa and CPSF 100 kDa are concentrated in nuclear domains closely

associated with coiled bodies and newly synthesized RNA. *EMBO J* **15**, 2883-92.

- Schultz, A., Nottrott, S., Watkins, N. J. & Luhrmann, R. (2006). Protein-protein and protein-RNA contacts both contribute to the 15.5K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs. *Mol Cell Biol* **26**, 5146-54.
- Schwarzacher, H. G. & Mosgoeller, W. (2000). Ribosome biogenesis in man: current views on nucleolar structures and function. *Cytogenet Cell Genet* **91**, 243-52.
- Segref, A., Mattaj, I. W. & Ohno, M. (2001). The evolutionarily conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export. *RNA* **7**, 351-60.
- Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U. & Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol* **8**, 27-31.
- Sen, G. L. & Blau, H. M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat Cell Biol* **7**, 633-6.
- Senger, B., Lafontaine, D. L., Graindorge, J. S., Gadai, O., Camasses, A., Sanni, A., Garnier, J. M., Breitenbach, M., Hurt, E. & Fasiolo, F. (2001). The nucle(ol)ar Tif6p and Efl1p are required for a late cytoplasmic step of ribosome synthesis. *Mol Cell* **8**, 1363-73.
- Sharma, K. & Tollervey, D. (1999). Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. *Mol Cell Biol* **19**, 6012-9.

- Shpargel, K. B. & Matera, A. G. (2005). Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. *Proc Natl Acad Sci U S A* **102**, 17372-7.
- Sienna, N., Larson, D. E. & Sells, B. H. (1996). Altered subcellular distribution of U3 snRNA in response to serum in mouse fibroblasts. *Exp Cell Res* **227**, 98-105.
- Singh, R. & Reddy, R. (1989). Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci U S A* **86**, 8280-3.
- Sirri V, Urcuqui-Inchima S, Roussel P, Hernandez-Verdun D (2008) Nucleolus: the fascinating nuclear body. *Histochemistry and cell biology* **129**: 13-31
- Sleeman, J. E. & Lamond, A. I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr Biol* **9**, 1065-74.
- Sleeman, J. E. & Lamond, A. I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr Biol* **9**, 1065-74.
- Snaar, S., Wiesmeijer, K., Jochemsen, A. G., Tanke, H. J. & Dirks, R. W. (2000). Mutational analysis of fibrillarin and its mobility in living human cells. *J Cell Biol* **151**, 653-62.
- Soler-Botija, C., Cusco, I., Caselles, L., Lopez, E., Baiget, M. & Tizzano, E. F. (2005). Implication of fetal SMN2 expression in type I SMA pathogenesis: protection or pathological gain of function? *J Neuropathol Exp Neurol* **64**, 215-23.
- Sontheimer, E. J. & Steitz, J. A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989-96.

- Sorger, P. K. & Pelham, H. R. (1987). Cloning and expression of a gene encoding hsc73, the major hsp70-like protein in unstressed rat cells. *EMBO J* **6**, 993-8.
- Spector, D. L. (1993). Macromolecular domains within the cell nucleus. *Annu Rev Cell Biol* **9**, 265-315.
- Spector, D. L. (2001). Nuclear domains. *J Cell Sci* **114**, 2891-3.
- Spector, D. L. (2006). SnapShot: Cellular bodies. *Cell* **127**, 1071.
- Stanek, D., Rader, S. D., Klingauf, M. & Neugebauer, K. M. (2003). Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J Cell Biol* **160**, 505-16.
- Sun, Y., Grimmmler, M., Schwarzer, V., Schoenen, F., Fischer, U. & Wirth, B. (2005). Molecular and functional analysis of intragenic SMN1 mutations in patients with spinal muscular atrophy. *Hum Mutat* **25**, 64-71.
- Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J. & Huttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci U S A* **99**, 7536-41.
- Terns, M. P. & Dahlberg, J. E. (1994). Retention and 5' cap trimethylation of U3 snRNA in the nucleus. *Science* **264**, 959-61.
- Terns, M. P., Grimm, C., Lund, E. & Dahlberg, J. E. (1995). A common maturation pathway for small nucleolar RNAs. *EMBO J* **14**, 4860-71.
- Terns, M. P. & Terns, R. M. (2002). Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr* **10**, 17-39.

- Tollervey, D. (1987). A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J* **6**, 4169-75.
- Tollervey, D. & Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Curr Opin Cell Biol* **9**, 337-42.
- Tollervey, D., Lehtonen, H., Jansen, R., Kern, H. & Hurt, E. C. (1993). Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell* **72**, 443-57.
- Torchet, C., Bousquet-Antonelli, C., Milligan, L., Thompson, E., Kufel, J. & Tollervey, D. (2002). Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs. *Mol Cell* **9**, 1285-96.
- Trainor, P. A., Dixon, J. & Dixon, M. J. (2009). Treacher Collins syndrome: etiology, pathogenesis and prevention. *Eur J Hum Genet* **17**, 275-83.
- Tran, E. J., Zhang, X. & Maxwell, E. S. (2003). Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J* **22**, 3930-40.
- Traub, P. & Nomura, M. (1968). Structure and function of *E. coli* ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc Natl Acad Sci U S A* **59**, 777-84.
- Tschochner, H. & Hurt, E. (2003). Pre-ribosomes on the road from the nucleolus to the cytoplasm. *Trends Cell Biol* **13**, 255-63.
- Tyc, K. & Steitz, J. A. (1989). U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J* **8**, 3113-9.

- Tycowski, K. T., Shu, M. D. & Steitz, J. A. (1994). Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. *Science* **266**, 1558-61.
- Tycowski, K. T., You, Z. H., Graham, P. J. & Steitz, J. A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol Cell* **2**, 629-38.
- Udem, S. A. & Warner, J. R. (1973). The cytoplasmic maturation of a ribosomal precursor ribonucleic acid in yeast. *J Biol Chem* **248**, 1412-6.
- Urlaub, H., Raker, V. A., Kostka, S. & Luhrmann, R. (2001). Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J* **20**, 187-96.
- Venema, J. & Tollervey, D. (1995). Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*. *Yeast* **11**, 1629-50.
- Venema, J. & Tollervey, D. (1999). Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet* **33**, 261-311.
- Venema, J., Vos, H. R., Faber, A. W., van Venrooij, W. J. & Raue, H. A. (2000). Yeast Rrp9p is an evolutionarily conserved U3 snoRNP protein essential for early pre-rRNA processing cleavages and requires box C for its association. *RNA* **6**, 1660-71.
- Verheggen, C., Lafontaine, D. L., Samarsky, D., Mouaikel, J., Blanchard, J. M., Bordonne, R. & Bertrand, E. (2002). Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J* **21**, 2736-45.
- Verheggen, C., Mouaikel, J., Thiry, M., Blanchard, J. M., Tollervey, D., Bordonne, R., Lafontaine, D. L. & Bertrand, E. (2001). Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain. *EMBO J* **20**, 5480-90.

- Vetter, I. R., Nowak, C., Nishimoto, T., Kuhlmann, J. & Wittinghofer, A. (1999). Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature* **398**, 39-46.
- Vidal, V. P., Verdone, L., Mayes, A. E. & Beggs, J. D. (1999). Characterization of U6 snRNA-protein interactions. *RNA* **5**, 1470-81.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. & Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell* **6**, 1331-42.
- Villa, T., Ceradini, F. & Bozzoni, I. (2000). Identification of a novel element required for processing of intron-encoded box C/D small nucleolar RNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* **20**, 1311-20.
- Villa, T., Ceradini, F., Presutti, C. & Bozzoni, I. (1998). Processing of the intron-encoded U18 small nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exo- and endonucleolytic activities. *Mol Cell Biol* **18**, 3376-83.
- Visintin, R. & Amon, A. (2000). The nucleolus: the magician's hat for cell cycle tricks. *Curr Opin Cell Biol* **12**, 752.
- Vithana, E. N., Abu-Safieh, L., Allen, M. J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N. D., Willis, C., Moore, A. T., Bird, A. C., Hunt, D. M. & Bhattacharya, S. S. (2001). A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell* **8**, 375-81.
- Wang, C. & Meier, U. T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J* **23**, 1857-67.

- Watkins, N. J., Dickmanns, A. & Luhrmann, R. (2002). Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol* **22**, 8342-52.
- Watkins, N. J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M. & Luhrmann, R. (1998). Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* **4**, 1549-68.
- Watkins, N. J., Lemm, I., Ingelfinger, D., Schneider, C., Hossbach, M., Urlaub, H. & Luhrmann, R. (2004). Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol Cell* **16**, 789-98.
- Watkins, N. J., Lemm, I. & Luhrmann, R. (2007). Involvement of nuclear import and export factors in U8 box C/D snoRNP biogenesis. *Mol Cell Biol* **27**, 7018-27.
- Watkins, N. J., Newman, D. R., Kuhn, J. F. & Maxwell, E. S. (1998). *In vitro* assembly of the mouse U14 snoRNP core complex and identification of a 65-kDa box C/D-binding protein. *RNA* **4**, 582-93.
- Watkins, N. J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C. & Luhrmann, R. (2000). A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**, 457-66.
- Wegierski, T., Billy, E., Nasr, F. & Filipowicz, W. (2001). Bms1p, a G-domain-containing protein, associates with Rcl1p and is required for 18S rRNA biogenesis in yeast. *RNA* **7**, 1254-67.

- Wehner, K. A., Ayala, L., Kim, Y., Young, P. J., Hosler, B. A., Lorson, C. L., Baserga, S. J. & Francis, J. W. (2002). Survival motor neuron protein in the nucleolus of mammalian neurons. *Brain Res* **945**, 160-73.
- Whitehead, S. E., Jones, K. W., Zhang, X., Cheng, X., Terns, R. M. & Terns, M. P. (2002). Determinants of the interaction of the spinal muscular atrophy disease protein SMN with the dimethylarginine-modified box H/ACA small nucleolar ribonucleoprotein GAR1. *J Biol Chem* **277**, 48087-93.
- Wieben, E. D., Nenninger, J. M. & Pederson, T. (1985). Ribonucleoprotein organization of eukaryotic RNA. XXXII. U2 small nuclear RNA precursors and their accurate 3' processing *in vitro* as ribonucleoprotein particles. *J Mol Biol* **183**, 69-78.
- Will, C. L. & Luhrmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol* **9**, 320-8.
- Will, C. L. & Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* **13**, 290-301.
- Winkler, C., Eggert, C., Gradl, D., Meister, G., Giegerich, M., Wedlich, D., Lagerbauer, B. & Fischer, U. (2005). Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. *Genes Dev* **19**, 2320-30.
- Wirth, B. (2000). An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). *Hum Mutat* **15**, 228-37.
- Wirth, B., Brichta, L., Schrank, B., Lochmuller, H., Blick, S., Baasner, A. & Heller, R. (2006). Mildly affected patients with spinal muscular atrophy are partially protected by an increased SMN2 copy number. *Hum Genet* **119**, 422-8.

- Wirth, B., Herz, M., Wetter, A., Moskau, S., Hahnen, E., Rudnik-Schoneborn, S., Wienker, T. & Zerres, K. (1999). Quantitative analysis of survival motor neuron copies: identification of subtle SMN1 mutations in patients with spinal muscular atrophy, genotype-phenotype correlation, and implications for genetic counseling. *Am J Hum Genet* **64**, 1340-56.
- Wohlwend, D., Strasser, A., Dickmanns, A. & Ficner, R. (2007). Structural basis for RanGTP independent entry of spliceosomal U snRNPs into the nucleus. *J Mol Biol* **374**, 1129-38.
- Wolin, S. L. & Cedervall, T. (2002). The La protein. *Annu Rev Biochem* **71**, 375-403.
- Xue, D., Robinson, D. A., Pannone, B. K., Yoo, C. J. & Wolin, S. L. (2000). U snRNP assembly in yeast involves the La protein. *EMBO J* **19**, 1650-60.
- Yang, Y., Isaac, C., Wang, C., Dragon, F., Pogacic, V. & Meier, U. T. (2000). Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140. *Mol Biol Cell* **11**, 567-77.
- Yean, S. L., Wuenschell, G., Termini, J. & Lin, R. J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881-4.
- Yong, J., Golembe, T. J., Battle, D. J., Pellizzoni, L. & Dreyfuss, G. (2004). snRNAs contain specific SMN-binding domains that are essential for snRNP assembly. *Mol Cell Biol* **24**, 2747-56.
- Yong, J., Pellizzoni, L. & Dreyfuss, G. (2002). Sequence-specific interaction of U1 snRNA with the SMN complex. *EMBO J* **21**, 1188-96.

- Young, P. J., Le, T. T., Dunckley, M., Nguyen, T. M., Burghes, A. H. & Morris, G. E. (2001). Nuclear gems and Cajal (coiled) bodies in fetal tissues: nucleolar distribution of the spinal muscular atrophy protein, SMN. *Exp Cell Res* **265**, 252-61.
- Yuan, G., Klambt, C., Bachellerie, J. P., Brosius, J. & Huttenhofer, A. (2003). RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res* **31**, 2495-507.
- Zebarjadian, Y., King, T., Fournier, M. J., Clarke, L. & Carbon, J. (1999). Point mutations in yeast CBF5 can abolish *in vivo* pseudouridylation of rRNA. *Mol Cell Biol* **19**, 7461-72.
- Zhang, H. L., Pan, F., Hong, D., Shenoy, S. M., Singer, R. H. & Bassell, G. J. (2003). Active transport of the survival motor neuron protein and the role of exon-7 in cytoplasmic localization. *J Neurosci* **23**, 6627-37.
- Zhang, X., Champion, E. A., Tran, E. J., Brown, B. A., 2nd, Baserga, S. J. & Maxwell, E. S. (2006). The coiled-coil domain of the Nop56/58 core protein is dispensable for sRNP assembly but is critical for archaeal box C/D sRNP-guided nucleotide methylation. *RNA* **12**, 1092-103.
- Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M. & Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133**, 585-600.
- Zhao, R., Kakiyama, Y., Gribun, A., Huen, J., Yang, G., Khanna, M., Costanzo, M., Brost, R. L., Boone, C., Hughes, T. R., Yip, C. M. & Houry, W. A. (2008). Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *J Cell Biol* **180**, 563-78.

- Zhao, X. & Yu, Y. T. (2004). Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in *Xenopus* oocytes. *RNA* **10**, 681-90.

Publications and presentations

Publications

McKeegan, K. S., Debieux, C. M., Boulon, S., Bertrand, E. & Watkins, N. J. (2007). A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol Cell Biol* **27**, 6782-93.

McKeegan, K.S., Debieux, C.M., & Watkins, N.J. (2009) The AAA+ proteins TIP48 and TIP49 regulate interactions between the core box C/D proteins and biogenesis factors during snoRNP formation. (in preparation)

Presentations

Oral presentation, RNA UK conference, Lake district, UK, Jan 18-20th 2008. “Involvement of the SMN complex in box C/D snoRNP biogenesis”

Poster presentation, Dynamic organisation of the nucleolus meeting, CSH, New York. Sept 17-21st 2008. “Involvement of the SMN complex in box C/D snoRNP biogenesis”

Awards

1st prize for oral presentation at RNA UK conference, Lake district, UK, Jan 18-20th 2008. “Involvement of the SMN complex in box C/D snoRNP biogenesis”

INVOLVEMENT OF THE SMN COMPLEX IN BOX C/D snoRNP BIOGENESIS

Charles M. Debieux, Kenneth S. McKeegan, Ira Lemm¹, Reinhard Lührmann¹, Nicholas J. Watkins. ICaMB, Newcastle University, UK. ¹ MPI-BPC, Gottingen, Germany.

In the eukaryotic nucleolus, the box C/D small nucleolar RNPs (snoRNPs) are essential for the processing and modification of the ribosomal RNA. Box C/D snoRNPs either direct 2'-O-methylation of the rRNA or function as RNA chaperones in pre-rRNA

processing. We have demonstrated that box C/D snoRNP biogenesis is a highly complex and co-ordinated process that is mediated by a large, dynamic pre-snoRNP complex in the nucleoplasm¹. This complex contains core box C/D proteins (15.5K, NOP56, NOP58, fibrillarin) and factors linked to assembly (TIP48, TIP49, NUFIP, NOP17, BCD1, TAF9), 3' processing (La, LSm, and exosome) and nuclear transport (Snurportin1, PHAX, Ran and CRM1)^{1,2,3}. SMN, the product of the gene linked to the neurodegenerative disorder spinal muscular atrophy, binds the core protein fibrillarin and has been linked to box C/D snoRNP biogenesis. However, this protein was not found in the pre-snoRNP complex¹. Indeed, SMN and the associated gemin proteins have so far only been shown to be important for the assembly of spliceosomal snRNPs.

We have investigated the potential role of SMN in box C/D snoRNP biogenesis. We show that SMN specifically interacts with the snoRNP assembly factors NUFIP and BCD1 *in vitro*. This suggests that the SMN complex interacts with the pre-snoRNP complex during the biogenesis process. Using RNAi, we found that the loss of either SMN or each of the individual gemin proteins did not significantly alter the levels of either the box C/D snoRNA or U1 snRNA in the cell. The exception was gemin5 which, when depleted, resulted in a decrease in the levels of the U3, U8 and U14 box C/D snoRNAs. The depletion of gemin5 also resulted in the reduction of U1 snRNA levels but had no effect on the amount of 7SL in the cell. Analysis of U3 and U8 snoRNA localisation in cells depleted of either SMN or the gemin proteins revealed that these factors are essential for correct localisation of the snoRNAs. Cells depleted of SMN or gemin proteins showed increased levels of the U3 and U8 snoRNAs in the nucleoplasm, the remnants of Cajal bodies and in some instances in the cytoplasm. We suggest that SMN, which likely functions through a transient interaction with the pre-snoRNP, is important for the efficient localisation of the snoRNP complexes.

- 1) Watkins, NJ, *et al.*, (2004) *Mol Cell.* 16, 789-98.
- 2) McKeegan, KS, *et al.*, (2007) *Mol. Cell. Biol.*, 27, 6782-93.
- 3) Watkins, NJ, *et al.*, (2007) *Mol. Cell. Biol.*, 27, 7018-27.